Solution NMR Studies on Fragments of the Yeast GPCR Ste2p

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. Nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Martin Poms

aus

Österreich

Promotionskomitee

Prof. Dr. Oliver Zerbe (Leitung der Dissertation)

Prof. Dr. Andreas Plückthun

Prof. Dr. Roland Riek

Zürich, 2015

Erklärung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe im Sinne der Promotionsordnung vom 03. Februar 2005 angefertigt. Bei der Abfassung der Dissertation wurden keine anderen als die darin angegebenen Hilfsmittel benutzt.

Zürich, Juli 2015

Martin Poms

Acknowledgements

First and foremost, I would like to express my special appreciation and gratefulness to my advisor Professor Dr. Oliver Zerbe for giving me this opportunity and for his continued support and guidance. I am also grateful for the other members of my thesis committee Prof. Dr. Andreas Plückthun and Prof. Dr Roland Riek for their time and helpful suggestions.

I want to thank all the members of the Zerbe group for providing scientific input, guidance and help and especially for creating such an exceptional work environment. I want to specifically extend my gratitude to Randall Watson, Christina Ewald and Jacopo Marino for welcoming me to Zürich (and the downstairs office) and producing the best coffee on Irchel campus. Your fruitful discussions about science, music and everything else under the sun as well as your misunderstandings were an intellectual treat. Furthermore, I would like to especially thank Martin Christen for his knowledge in all matters NMR, Westeros and science fiction, and Philipp Ansorge for delivering constant happiness, beef and barbeque.

I want to thank Prof. Dr. Fred Naider and his group for the effective collaboration and their knowledge in peptide chemistry, especially Katrina Fracchiolla and Leah Cohen, who have been great colleagues across the Atlantic.

I would also like to thank Peter Güntert and Daniel Gottstein for their help in structure calculations, Gunnar Jeschke and Simon Böhm for their equipment and help with EPR measurements, as well as the Robinson group for sharing their equipment and the organization of intergroup activities.

For their constant support in all matters NMR and the maintenance of the NMR facility, I would like to thank Nadja Bross and Simon Jurt, whose assistance in all things concerning NMR and computers has been invaluable.

Finally, I want to thank my family, whose long-distance support has been priceless, especially my parents and my brother Christopher for balancing science with philology and for all the times he made me laugh.

I could not have done it without you. Let's see what's next...

Summary

Membrane proteins have become one of the most interesting classes of biomolecules for structural studies. Located in membranes, they are usually involved in the transfer of information and substances between cells and their environment and are therefore highly coveted biochemical targets. One specific class of membrane proteins are G-protein coupled receptors (GPCRs). GPCRs exist in all eukaryotic organisms and are involved in a plethora of physiological processes. They function as receivers for extracellular stimuli and are responsible for the signal transduction into the cell. GPCRs respond to a multitude of different external signals, such as photons, ions, small organic molecules and fatty acids, peptides, hormones and even entire proteins. Therefore, they play a key role in visual, gustatory and olfactory detection, behavioral and mood regulation, inflammatory response and the regulation of sympathetic as well as parasympathetic nervous system. The human genome contains information for approximately 800 different GPCRs, and because of their involvement in signaling it is not of much surprise that they present the target of almost 40% of drugs currently on the market or in clinical trials.

In spite of their highly diverse functions, all GPCRs exhibit a very conserved overall structure. They are constructed by a single polypeptide chain that folds into seven alpha-helical transmembrane helices (TMs), which are connected through three intracellular (ICL) and three extracellular loops (ECL). Upon interaction with an extracellular ligand (by binding the ECLs and potentially interacting with the helices themselves), structural changes propagate throughout the transmembrane region. This results in conformational changes in the intracellular region, which in turn allows for the interaction with downstream effectors. These downstream effectors, most notably heterotrimeric G-proteins and arrestins, thus trigger a cascade of downstream signaling events.

The knowledge of the exact atomic structure of GPCRs therefore permits the design of drugs specifically inhibiting or facilitating signal transduction. However, to date only a handful of high-resolution GPCR structures has been solved. The reason for this can be found in several inherent unfavorable properties, such as their low natural abundance and expression levels, instability, hydrophobic nature and inherent flexibility.

In this work, we investigated methods and strategies to counter these problems specifically associated with structural studies of GPCRs, using a divide-and-conquer approach by splitting the

yeast Ste2p receptor into fragments containing several transmembrane helices, and studying them using solution NMR spectroscopy. In Chapter 1 general problems encompassed in solution structure determination of membrane proteins and GPCRs specifically will be discussed in detail. Strategies for recombinant expression and purification of membrane proteins, specifically GPCR fragments will be introduced, as well as strategies for the identification of a suitable membrane mimetic. Furthermore, solution NMR spectroscopy as a viable option for the structure determination of membrane proteins will be discussed. The feasibility of different NMR techniques and biochemical methods, such as different isotope labeling schemes, and approaches for structure determination of membrane proteins using sparse NMR data will also be debated.

Chapters 2 and 3 will present NMR and biophysical analysis of several fragments from the yeast GPCR Ste2p, containing up to three transmembrane helices. TM1 (containing the first transmembrane helix), TM12, TM123 and TM127, a construct in which TM7 is covalently linked to TM2 were expressed as inclusion bodies in *E. coli*, either as fusion proteins with truncated N-termini using the ΔTrp leader sequence as a signaling peptide, or via direct expression containing the full N-terminus. The fragments have been characterized using various NMR experiments in detergent micelles or in organic solvent mixtures, focusing on the identification of interhelical contacts.

Another aspect when investigating GPCR fragments is the study of receptor folding. It is widely believed, that receptor biosynthesis follows a two-stage model. First, the individual helices are synthesized and released into the membrane interior via the translocon. Subsequently, folding of the individual helices occurs spontaneously, as secondary structure of membrane proteins, in contrast to soluble proteins, is mainly dependent on the amino acid sequence and the respective propensity to form α-helices, and their hydrophobic exterior (compared to a distinct hydrophobic core in soluble proteins). Then, they diffuse through the the membrane until native contacts are formed and the mature receptor is assembled. Therefore, systematic elongation of receptor fragments by subsequent addition of further helices permits the investigation of the formation of interhelical contacts during early receptor folding.

In our study, we see that all constructs are integrated well into the micelle, and that TM1, which in isolation is destabilized, is rigidified by the addition of an additional helix in the previously solved TM12 fragment to form a stable hairpin. Further elongation to the TM123 construct however, results in several unsatisfied interhelical contacts by polar residues located in the

center of TM3. This in turn leads to a high propensity of TM123 to aggregate, excluding the use of more detailed NMR experiments, which are needed for a high-resolution structure. Therefore, a construct in which TM3 is substituted by TM7 - as based on a homology model, it is believed that most interhelical contacts form between TM1, TM2 and TM7 - was designed, yielding the fragment TM127. TM127 proves to be much more stable than TM123, but exhibits more interhelical dynamics than the rather stable TM12 fragment. Our data indicate that in case of the TM127 construct, TM2 and TM7 compete for the same interhelical contacts to be formed with TM1. It has to be noted, that the assignment of longer fragments was greatly aided by knowledge of their approximate chemical shifts from the already assigned shorter constructs, at least for overlapping parts. Close contacts of amide protons, visible in the ^{15}N-resolved NOESY, will almost exclusively be observed to sequential protons, and are less likely influenced by changes in tertiary structure. Strips in 3-dimensional NOESY spectra from different size fragments with sequential overlap can therefore be compared efficiently, and resonance assignments can be adapted through a strip match. We were able to automate this process and successfully transfer over 90% of assignments from the TM12 to the TM127 fragment, for example.

High-resolution structures of several fragments have been determined using NMR spectroscopy in this work. The structure of TM123 was determined in a trifluoroethanol/water mixture and while the three helices were properly formed, tertiary contacts could not be observed. The structure of TM1 was solved in detergent micelles, showing the formation of a helix, which is destabilized around a G-XXX-G pentapeptide stretch, known to be a common motif for interhelical contacts. These contacts are satisfied once the second helix (TM2) is added, as was shown in the previously solved structure of TM12 in detergent micelles. As was mentioned before, TM123 proved to be too unstable for structure determination, and NMR studies have indeed revealed destabilization around polar residues in the center of TM3, which do not properly insert into the micelle, possibly being the cause of the observed aggregation. Membrane proteins often do not exhibit enough long-range contacts for complete structure elucidation, which is why other NMR experiments utilizing sparse data are often employed. One example of sparse NMR data are paramagnetic relaxation enhancement (PRE) measurements. For this purpose, paramagnetic spinlabels are attached to specific sites in the protein, effectively bleaching signals in spatial proximity, at much longer distances than conventional experiments. Furthermore, residual dipolar coupling (RDC) experiments allow the determination of the absolute orientation of specific chemical bonds, relative to the overall protein alignment orientation. While at this point we could not determine a unique structure of TM127 in micelles, probably due to high flexibility of the

system, sparse sampling data experiments have allowed us to characterize overall topological features. Even though TM127 appears to be a highly dynamic system and does not present a rigid helix bundle, we could show that the helices are in place, exhibit at least transient contacts and are firmly integrated into the micelle.

The implications of those findings for further work of our group on the study of membrane proteins, in particular GPCRs, will be reviewed in Chapter 4. Apart from strategies for the expression and purification of membrane protein fragments, their usefulness as an instrument to study interhelical contacts and more importantly GPCR folding will be discussed. Experiments to probe for conformational preferences of GPCR fragments expressed directly in the *E. coli* inner membrane, using reporter assays through fusion proteins, will be highlighted. Furthermore, Chapter 4 will introduce another concept for the study of extracellular loops of GPCRs, by grafting them onto more rigid, hence more stable beta-barrel proteins, and to examine the ligand binding via chemical shift mapping.

Zusammenfassung

In der Strukturbiologie sind Membranproteine in letzter Zeit zu einer der interessantesten Klassen von Biomolekülen avanciert. Da sie, wie der Name erahnen lässt, in Membranen vorkommen, besteht ihre Hauptaufgabe im Transfer von Molekülen und Information zwischen Zellen und deren Umwelt, weshalb sie auch ein äusserst begehrtes biochemisches Ziel darstellen. Eine bestimmte Gruppe von Membranproteinen sind die 'G-Protein gekoppelten Rezeptoren' (GPCRs – *engl. G-protein coupled receptors*). GPCRs sind in allen eukaryontischen Organismen vorhanden und bestimmen eine Vielzahl an physiologischen Prozessen. Sie dienen als Empfänger für extrazelluläre Stimuli und sind für deren Signalübertragung ins Zellinnere verantwortlich. GPCRs werden von einer Vielzahl an unterschiedlichen externen Signalen angesprochen, wie zum Beispiel Photonen, Ionen, kleinen organischen Molekülen und Fettsäuren, Peptiden, Hormonen und sogar vollständigen Proteinen. Deshalb spielen sie auch eine Schlüsselrolle in der Detektion von visuellen, geschmacklichen und olfaktorischen Reizen, Gefühls- und Verhaltenssteuerung, Entzündungsreaktionen, sowie der Regulation des sympathischen als auch parasympathischen Nervensystems. Das menschliche Genom enthält die Information für ungefähr 800 verschiedene GPCRs, und aufgrund ihrer prominenten Rolle in Signalvorgängen ist es keine Überraschung, dass sie zum jetzigen Stand das Ziel von fast 40% aller am Markt erhältlichen oder gerade in klinischen Studien getesteten Medikamente sind.

Trotz ihrer höchst diversen Funktionen besitzen GPCRs jedoch eine sehr konservierte Grundstruktur. Sie bestehen aus einer einzelnen Polypeptidkette, die sich in sieben alpha-helikale Transmembranhelices (TMs) faltet. Diese sind wiederum durch drei intrazelluläre (ICLs) und drei extrazelluläre (ECLs) 'loops' verbunden. Bei der Interaktion mit einem extrazellulären Liganden, der durch die ECLs gebunden wird und potentiell auch mit den Helices selbst interagiert, werden strukturelle Änderungen in der Transmembranregion hervorgerufen. Diese resultieren in konformationellen Änderungen in der intrazellulären Region, was wiederum die Interaktion mit nachgeschalteten Effektoren bewirkt. Diese Effektoren, hauptsächlich heterotrimere G-Proteine und Arrestin, lösen dann eine Signalkaskade im Zellinneren aus.

Die Kenntnis der exakten atomaren Strukturen von GPCRs erlaubt daher die Gestaltung von Medikamenten, welche die Signalübertragung spezifisch inhibieren oder verstärken. Dennoch sind bis heute nur eine Handvoll an hochauflösenden GPCR Strukturen bestimmt worden. Als Grund hierfür gelten einige ungünstige, inhärente Eigenschaften von GPCRs, so wie ihr geringes

natürliches Vorkommen, die geringen Expressionsausbeuten, Instabilität, Hydrophobizität und Flexibilität.

In dieser Arbeit werden Methoden und Strategien untersucht, diesen mit der Strukturaufklärung von GPCR spezifischen einhergehenden Problemen entgegenzuwirken. Hierfür wurde eine 'divide-and-conquer' Strategie entwickelt, bei welcher der Hefe Rezeptor Ste2p in Fragmente mit einer oder mehreren Transmembranhelices aufgeteilt und mittels Lösungs-NMR Spektroskopie untersucht wird. Im ersten Kapitel werden die spezifischen Probleme, die mit der Strukturaufklärung von Membranproteinen und GPCRs im Speziellen in löslicher Umgebung einhergehen, im Detail diskutiert. In diesem Rahmen werden Strategien für die rekombinante Expression und Aufreinigung von Membranproteinen und im speziellen GPCR-Fragmenten, sowie Strategien für die Bestimmung eines geeigneten Membranmimetikums, vorgestellt. Ausserdem wird die Tauglichkeit von verschiedenen NMR Techniken und biochemischen Methoden, wie zum Beispiel verschiedener Isotopenmarkierungsmodelle und Vorgehensweisen zur Strukturaufklärung von Membranproteinen, bei denen traditionell nur relativ wenige NMR Restraints ermittelt werden, erörtert.

In Kapitel 2 und 3 wird die Analyse von mehreren Fragmenten des Hefe GPCRs Ste2p mit bis zu drei Transmembranhelices mittels NMR und biophysikalischen Methoden präsentiert. TM1 (enthält die erste Transmembranhelix), TM12, TM123 und TM127 (ein Konstrukt, in dem TM7 kovalent mit TM2 verbunden wurde) wurden als Einschlusskörperchen in *E. coli*, entweder als Fusionsprotein mit gekürztem N-terminus und der ΔTrp Signalsequenz oder direkt mit vollständigem N-terminus, exprimiert. Die Fragmente wurden mithilfe von unterschiedlichen NMR Experimenten in Detergenzmizellen oder in Mischungen mit organischen Lösungsmitteln, mit Fokus auf der Identifikation von interhelikalen Kontakten, untersucht.

Ein weiterer wichtiger Aspekt bei der Untersuchung von GPCR-Fragmenten ist die Erforschung der Faltung von Rezeptoren. Die vorherrschende Meinung ist, dass die Rezeptorbiosynthese einem zweistufigen Modell folgt. Als erstes werden die einzelnen helikalen Peptidsegmente über das Translocon in die Membran eingespeist, da die Sekundärstruktur von Membranproteinen im Gegensatz zu löslichen Proteinen hauptsächlich von der Aminosäuresequenz und deren entsprechender Neigung zur Bildung von α-Helices, sowie von deren hydrophober Aussenseite (im Kontrast zum hydrophoben Kern in löslichen Proteinen) definiert wird. Dort wird die helikale Sekundärstruktur spontan gebildet. Anschliessend diffundieren die einzelnen Helices

durch die Membran bis sich native interhelikale Kontakte gebildet haben und schließlich der gesamte Rezeptor aufgebaut ist. Die systematische Erweiterung von Rezeptorfragmenten durch die Addition weiterer Helices ist daher gut geeignet, interhelikale Kontakte im Frühstadium der Rezeptorfaltung zu untersuchen.

In unseren Studien sehen wir, dass alle Konstrukte gut in die Mizelle integrieren und dass TM1, welches isoliert destabilisiert ist, durch die Addition einer weiteren Helix stabilisiert wird, wie in der bereits gelösten Struktur des TM12 Fragmentes, das einen stabilen 'hairpin' formt, zu sehen ist. Eine weitere Verlängerung zum TM123 Fragment resultierte jedoch in einigen unbefriedigten polaren Resten in der Mitte von TM3. Dies wiederum führte zu einer hohen Neigung von TM123 zu aggregieren, wodurch es nicht möglich war, komplexere und zeitintensivere NMR Experimente, die für die Strukturaufklärung nötig wären, durchzuführen. Deswegen wurde ein weiteres Fragment entworfen, TM127, in dem TM3 durch TM7 ersetzt wurde, da, basierend auf einem Homologiemodell, die meisten interhelikalen Kontakte zwischen TM1, TM2 und TM7 vermutet werden. Obwohl TM127 sehr viel stabiler als TM123 ist, zeigte es mehr interhelikale Dynamik als das eher rigide TM12. Unsere Daten weisen darauf hin, dass TM7 mit denselben interhelikalen Kontaktstellen an TM1 konkurriert wie TM12. Ausserdem muss erwähnt werden, dass die Zuordnung der chemischen Verschiebungen in NMR Spektren von längeren Fragmenten durch die Kenntnis ihrer ungefähren Position in überlappenden Teilen aus kürzeren Fragmenten sehr erleichtert wurde. Nahe Kontakte von Amidprotonen, die in ^{15}N-edititierten NOESY Spektren zu beobachten sind, werden fast ausschliesslich von sequentiellen Protonen beeinflusst, ohne nennenswerten Einfluss der Tertiärstruktur. Strips 3-dimensionaler NOESY Spektren von Fragmenten mit unterschiedlichen Grösse, aber überlappender Sequenz, können deshalb effizient miteinander verglichen und mittels 'strip matching' angepasst werden. Wir konnten diesen Prozess erfolgreich automatisieren und so zum Beispiel über 90% der Zuordnung von TM12 auf TM127 übertragen.

Ausserdem werden in dieser Arbeit hochauflösende Strukturen mehrerer Fragmente mittels NMR Spektroskopie gelöst. Die Struktur von TM123 wurde in einer Trifluorethanol/Wasser Mischung bestimmt. Während sich alle drei Helices richtig geformt haben, konnten keine tertiären Kontakte festgestellt werden. Die Struktur von TM1 wurde in Detergenzmizellen bestimmt, wobei wiederum die Bildung einer Helix zu sehen war, die jedoch im Bereich einer G-XXX-G Sequenz – ein bekanntes Motiv zur Bildung von interhelikalen Kontakten – destabilisiert war. Dieser Bereich wurde jedoch durch die Addition von TM2 gesättigt, wie in der bereits veröffentlichten Struktur von

TM12 zu sehen ist. Wie bereits erwähnt, entpuppte sich das TM123 Fragment als zu instabil für ausgiebige Strukturanalysen. NMR Untersuchungen zeigten in der Tat einige destabilisierte polare Reste im Zentrum von TM3, die nicht richtig in die Mizelle integriert werden und vermutlich der Grund für das hohe Aggregationspotential von TM123 sind. Membranproteine weisen häufig leider nicht genügend weitreichende Kontakte für eine umfassende Strukturaufklärung auf, weshalb oft NMR Experimente, die nur eine geringe Anzahl von NMR Daten liefern, angewendet werden. Eines dieser Experimente ist die paramagnetische Relaxationsverstärkung (PREs *engl. Paramagnetic relaxation enhancement*). Zu diesem Zweck wird ein paramagnetisches Spinlabel an einer bestimmten Position im Protein angebracht, das die Signale in räumlicher Nähe (aber auch in viel grösseren Distanzen als in herkömmlichen Experimenten) effektiv abschwächt. Im Weiteren können residuale dipolare Kopplungen (RDCs) Informationen über die absolute Orientierung von spezifischen chemischen Bindungen in Relation zur globalen Orientierung des Proteins liefern. Während wir zu diesem Zeitpunkt noch keine eindeutige Struktur von TM127 in Mizellen bestimmen konnten, was vermutlich auf die hohe Flexibilität des Systems zurückzuführen ist, konnten wir mithilfe der zuvor beschriebenen NMR Daten dessen grundlegende topologische Eigenschaften charakterisieren. Obwohl TM127 ein hochdynamisches System zu sein scheint und kein rigides Helixbündel aufweist, können wir zeigen, dass die Helices gebildet wurden, dabei fest in die Mizelle integriert waren und zumindest transiente Kontakte zueinander zeigten.

Die Implikationen dieser Erkenntnisse für weitere Projekte unserer Gruppe in der Untersuchung von Membranproteinen werden in Kapitel 4 diskutiert. Neben Strategien für die Expression und Aufreinigung von Membranproteinfragmenten wird deren Nützlichkeit zur Untersuchung von interhelikalen Kontakten und, noch wichtiger, zur GPCR Faltung diskutiert. Experimente zur Untersuchung von konformationellen Vorlieben von Membranproteinfragmenten durch die direkte Expression in die innere Membran von *E. coli* mithilfe von 'Reporter-assays' durch Fusionsproteine werden aufgezeigt. Ausserdem wird in Kapitel 4 ein weiteres Konzept zur Analyse der Ligandenbindung von extrazellulären loops von GPCRs mittels 'chemical shift mapping' vorgestellt, indem diese auf stabilere und daher rigidere beta-barrel Proteine transplantiert werden.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 GPCRs

G-protein coupled receptors (GPCRs) are membrane proteins involved in signaling processes and therefore constitute one of the most important classes of proteins in nature, as is evidenced by their abundance (3% of the human genome encode for GPCRs) and their diversity (over 2000 GPCRs are known in nature)[1, 2]. They function as receivers for extracellular stimuli and transduce signals to the inside of the cell. They interact with a bevy of different classes of ligands, such as light, peptides and even proteins, fatty acids, hormones, ions, as well as a multitude of small organic molecules and are involved in numerous physiological processes. Their importance is further confirmed by the fact that a multitude of currently available drugs target GPCRs[3, 4] and culminated in the bestowal of the Nobel prize in chemistry in 2012 to Brian Kobilka and Robert Leftkowitz for their work on GPCRs[5].

All GPCRs consist of a single polypeptide chain that folds into seven α-helical transmembrane helices (TMs) which are linked by three intracellular (IL) and three extracelluar (EL) loops, with an extracellular N-terminus and the C-terminus located in the cytosol[6]. While the transmembrane helix bundle displays a comparatively conserved overall fold among different receptors, individual helices can still exhibit significant structural differences. Their topology is determined by the existence and location of tertiary interhelical contacts, such as characteristic hydrophobic patterns and hydrogen bonds, as well as conserved signature motifs and conserved microswitches, which are imperative for function and activation, such as the G-XXX-G motif[7], or the D[E]RY[8, 9] and NPXXY motives[10]. As ligand binding induces structural changes in the TM-bundle, it is often quite flexible, with proline-induced kinks in helices[11], that are imperative for signal transduction, not uncommon. Despite relative low sequence homology between different receptors, the topology of the helix bundle remains comparatively conserved with the most notable differences developing more distinctly towards the extracellular domain.

The extracellular domain, containing the three extracellular loops as well as the N-terminus, is mainly responsible for ligand binding and therefore exhibits noticeable differences depending on the nature of the ligand. The mode of ligand binding can be highly diverse, as ligands can bind just to the extracellular part or be directed into the transmembrane region via a 'fly-casting' mechanism[12]. This diversity on one hand presents the basis for the categorization of different GPCR

classes[13] and on the other hand results in a high variation of arrangements and secondary structure elements.

While several classifications of GPCRs based on sequence and structural similarities have been proposed, the most commonly used groups GPCRs into six classes[13]: Class A or Rhodopsin-like, which contains 19 subgroups and presents the largest group, Class B (the Secretin receptor family), Class C (Metabotropic glutamate/pheromone), Class D (Fungal mating pheromone receptors), Class E (Cyclic AMP receptors) and Class F (Frizzled/Smoothened).

The primary function of the intracellular domain including the C-terminus, which often contains an eighth helix, is the signal transmission into the cell. It interacts with downstream effectors, most importantly G-protein heterotrimers and arrestins. Differences in the IC-domain therefore are comparably smaller than in other compartments of GPCRs and commonly affect selectivity and specificity of the signaling pathway or G-protein subtype preference. At the same time, intracellular domains undergo consequential conformational change upon receptor activation, which causes them to exhibit a high degree of flexibility in general[14]. As highly dynamic systems, GPCRs in nature (and to a lesser degree also *in vitro*) alternate between several conformations, their ground state, their excited state and possibly several intermediate states, even without a corresponding ligand, which results in basal activity for most receptors. One important property that has been discussed recently extensively is *biased signaling*. The latter refers to the fact that the same GPCR, depending on the exact nature of the ligand, can signal in different ways.

In spite of their pharmacological importance and recent advances in structure elucidation techniques, a limited number of high-resolution GPCR structures has been determined to date[9, 15-43]. The reason for this can be found in several characteristics of GPCRs. First of all, the hydrophobicity of GPCRs presents a major challenge. Naturally located in membranes, GPCRs require to be solubilized by membrane mimetics at all times, considerably complicating protein expression, purification and sample preparation in active form for structural studies. Another problem encountered when working with GPCRs is their inherent conformational flexibility, which is a necessary feature for effective signal transduction, but can interfere with the formation of well ordered crystals and cause significant line-broadening in NMR spectroscopy.

1.1.1 Expression and Purification

The first step in obtaining a sample feasible for structural studies obviously is the successful expression and purification of a GPCR. Even though the first structure of a GPCR, the structure of bovine rhodopsin[15], could be determined by extraction from natural sources (in this case retina of the bovine eye), GPCRs in general exhibit quite low natural abundance, demanding the implementation of recombinant expression techniques. Furthermore, over-expression of GPCRs in cell membranes frequently interferes with native cell signaling and consequently exhibits significant cell toxicity.

Cultured mammalian cells in general present the optimal expression system for mammalian GPCRs, as they possess the native membrane composition and provide the necessary machinery for correct post-translational modifications, which are known to stabilize GPCRs and play a role in their activity. While there have been some cases of successful cases[44-46], GPCR expression in mammalian cells proves to be extraordinarily cumbersome and relatively expensive and often does not produce sufficient amounts of protein needed for crystal screening or NMR spectroscopy. Furthermore, the introduction of isotope substitutions, which is necessary for some structural determination techniques, such as NMR, can be quite tedious and expensive.

The most common GPCR expression system for crystallography studies are insect cells[16, 19-21, 27, 28, 32, 35, 37-39, 47, 48]. While also time-consuming and quite costly, expression in insect cells often results in acceptable amounts of receptor (in the milligram per liter range) with a similar degree of post-translational modifications compared to mammalian cells. However, it may be necessary to supplement the growth medium with additional lipids to better mimic natural membrane composition. Moreover, isotopic labeling is not yet established in all cases. Additionally, expression in yeast cells, such as *Pichia pastoris*, proved to be a viable expression system due to it's relative simplicity to scale up, cost efficiency and possible isotope substitution as well as deuteration[49, 50].

Even though the *E. coli* expression system lacks the molecular machine to correctly process eukaryotic integral membrane proteins, it is one of the most efficient expression systems, due to the fast doubling time, inexpensive growth media that allows for easy isotope labeling, and most of all the relative effortlessness with which it can be manipulated genetically[51, 52]. Expression in *E. coli* often requires the implementation of a fusion protein to either direct expression into the inner membrane with cytoplasmic β-galactosidase[53], MBP[54, 55], TrxA[56] and Mistic[57], or into inclusion bodies, using GST[58] or TrpΔLE[59]. The utilization of a fusion protein commonly increases expression

yields, however, additional cleavage and purification steps might become necessary. Recently, another related expression system, cell-free protein expression[60], has become increasingly important. The easiest and most common application of cell-free expression is the utilization of *E. coli* extract, which allows for background-free expression, expression of toxic proteins and highly specified isotopic labeling schemes without isotope scrambling.

The purification of heterologously expressed GPCRs generally includes the solubilization in detergent, purification with chromatography and subsequent refolding of the receptor or isolation of a distinct (usually active) species.

1.1.2 Identifying a suitable membrane mimetic

Identifying the optimal solubilization agent for structural determination experiments can be rather challenging. As it is quite difficult to generally mimic natural membranes of different chemical composition, the right membrane mimetic needs to be determined on a case-by-case basis, depending on protein characteristics and limitations of the relevant structure determination technique employed. In general, a feasible membrane mimetic has to be chosen as a trade-off between two important factors: it needs to *i)* sufficiently solubilize the membrane protein to keep it stable and in solution, while at the same time *ii)* not being too harsh to denature the protein or impose artificial conformational constraints (reviewed here[61]).

There are two principal strategies that are employed during membrane protein preparation in regard to membrane model media, depending on expression and protein specific properties. The first approach is to directly express the properly folded protein in membranes and then solubilize it using a mild detergent that is suitable to extract the protein from the membrane, while not disrupting its native fold. The most prevalent detergent for this purpose is dodecylmaltoside (DDM), which has been shown to mostly preserve protein functionality[62, 63] which can usually be maintained throughout subsequent purification steps. However, direct expression into membranes is often accompanied by overall low yields, possible cell toxicity and the requirement for a signal sequence, which generally needs to be removed in an additional purification step, which in turn introduces another source of potential loss of protein.

Therefore, another strategy directing expression into inclusion bodies is often employed, generally leading to higher amounts of protein. As dense protein-lipid aggregates, inclusion bodies

most of the time can only be solubilized with the use of harsh detergents or in high concentration of chaotropic agents, such as urea or guanidinium hydrochloride, or even a combination of both[64-66]. During subsequent purification, the protein requires to be refolded by exchanging the membrane mimetic to non-denaturing detergents or detergent-lipid mixtures. Finding the right membrane mimetic to achieve proper refolding of the protein that is compatible with the desired application (this naturally also applies to direct membrane expression) can be quite challenging and often requires extensive optimization. The following subsections will highlight the advantages and disadvantages of the most prevalent membrane mimetics.

1.1.2.1 Micelles

Detergent micelles are the most common agent in membrane protein solubilization, purification and sample preparation. Micelles are formed when the concentration of detergents monomers in a solution reaches a detergent specific concentration, the critical micelle concentration (CMC), at which point aggregates with solvent exposed hydrophilic head groups on the outside and hydrophobic tails in the center are assembled. Dependent on alkyl chain length and bulkiness of the head group, detergent micelles can be found in oblate and prolate, and in very rare cases spherical forms[67]. Micelles can solubilize membrane proteins by assembling around the hydrophobic patches in a belt-like fashion and therefore can be utilized to either solubilize a membrane protein as a first step for purification or consequently as an environment for structure determination.

For structure determination studies it is often helpful to keep the protein-detergent complex relatively small to favor the formation of crystal contacts when employing x-ray crystallography or to improve tumbling rates in case of NMR spectroscopy. It has to be mentioned that the size of the detergent-protein complex depends largely on the surface area of the transmembrane domain of the protein, therefore introducing detergents that tend to form smaller micelles by themselves will not necessarily reduce the size of the detergent-protein complex (mixed micelles). Furthermore, solubility is not necessarily indication for the best possible detergent for structure determination, and thus detergent exchange between purification and measurement is frequently necessary.

While micelles do not play a big role in membrane protein structure determination using x-ray crystallography, they are still the most abundant membrane mimetic, when utilizing solution state NMR, mostly due to their relative small compared to other membrane-like systems. As stated above, the determination of the optimal detergent(s) for a specific membrane protein needs to be

established empirically in each individual case. While theoretically any detergent could be well suited for structure determination, a relatively short list of useful detergents for NMR-spectroscopy has emerged. Very harsh detergents, such as sodium dodecylsulfate (SDS) or lauryldimethylamine oxide (LDAO) often result in poor spectral quality, at least for α-helical membrane proteins, while alkyl glucosides as dodecyl dimaltoside (DDM), which are favored in x-ray crystallography, often do not prove powerful enough to efficiently solubilize membrane proteins for NMR spectroscopy. Therefore, the most common classes of detergents that yield reasonable spectra are alkyl phosphocholines, such as dodecyl phosphocholine (DPC), and lysophospholipids, such as palmitoyl lysophosphatidyl glycerol (LPPG) and myristyl lysophosphatidyl choline (LMPC), which have been successfully employed[66, 68, 69], sometimes as a mixture of two or more components[70]. Schematic representations for the most commonly used detergents can be found in Figure 1.1.

Micelles, however, are dynamic systems, in which detergent molecules from micelles are in constant exchange with residual free monomers in the solution. This might cause instability, if the detergent concentration is close to the CMC. Additionally, it is recommended to ensure a high ratio of empty micelles to reduce the chance of dimerization and aggregation. In fact, it has been shown that it is favorable to utilize detergent concentrations orders of magnitude above the CMC to improve sample stability[71].

1.1.2.2 Bicelles

Another useful system for the solubilization of membrane proteins are bicelles[72, 73]. Bicelles are aggregates of short stretches of lipidic bilayers that are stabilized by short-chain detergents and therefore perfectly simulate native membranes, while at the same time retaining relatively small sizes. The effective size of bicelles can be regulated by the ratio of the concentration of lipid to concentration of detergent. An accurate representation of this ratio is the q-factor, which can be calculated by dividing the total molarity of lipids by the total detergent concentration minus the CMC of the detergent to correct for the amount of free detergent in solution. For NMR studies, a q-factor of around 0.25 to 0.5 is desired, equating a relatively low lipid to high detergent ratio, resulting in stable and comparatively small bicelle sizes. Q-factors above 0.5 result in bicelles too large for practical NMR spetroscopy, while q-values under 0.25 lead to poorly defined membrane media, more similar to lipid doted micelles than stable bilayers. In NMR spectroscopy the most commonly employed system is a combination of dialkyl phosphatidylglycerol or dialkyl phosphatidylcholine, such as dimyristoyl phosphatidylcholine (DMPC) as a lipidic component and a

short-chain dialkyl phosphatidylcholine, such as dihexanoyl phosphatidylcholine (DHPC) as the detergent.

Thus, small isotropic bicelles present the most membrane-like environment that theoretically still allows solution-NMR structure determination, however, their application so far has been limited to very stable[74] or small membrane proteins[75, 76], or to the study of ligand interactions[77], not yet for complete structure determination of larger proteins. While several membrane protein structures in bicelles have been solved using solid-state NMR spectroscopy[51] and X-ray crystallography[78-80], no GPCRs have been crystalized from bicelles yet.

1.1.2.3 Nanodiscs and other membrane mimetics

Similar to bicelles, nanolipoproteins or nanodiscs stabilize the lipid bilayer by surrounding it with two molecules of a helical, amphiphatic protein, the membrane scaffold protein (MSP)[81, 82]. Their size still makes them useful for solution NMR studies, as their tumbling rate corresponds to a globular protein of roughly 200kDa[83]. It has been possible to functionally incorporate and study a variety of membrane proteins in nanodiscs[84, 85], including bacteriorhodopsin[86] and the β_2-adrenergenic receptor[87]. Similarly to nanodiscs, it is possible to stabilize the uni-lamellar lipid bilayer using copolymers instead of the protein belt, forming so-called lipodisqs[88, 89].

Another recent development has been the emergence of amphipols as an alternative to conventional detergents. Amphipols are polymers with alternating hydrophobic and hydrophilic elements that have been shown to stabilize internal membrane proteins, often preserving their functionality[90-92]. Furthermore, it has been shown that it is possible to refold GPCRs in an amphipol environment[93]. Amphipols theoretically can be used in combination with lipids or detergents, or just by themselves as the main solubilization agent. While NMR spectra of membrane proteins in amphipols have shown promising results[94], further characterization and optimization will be required.

Reverse micelles, in which detergents are dissolved in an organic solvent and therefore form inverse micelles with a 'water pocket' in the center have also drawn some interest as potential membrane mimetics. Even though the protein-detergent complex is potentially larger than in conventional micelles, the low viscosity of organic solvents and subsequent faster tumbling rate has

resulted in quite satisfying NMR spectra. In fact, in some cases the correlation time of membrane proteins has decreased by factor 3-6[95] in reversed micelles compared to conventional micelles.

Lipidic cubic phases (LCP) have also greatly aided in the crystallization of membrane proteins (reviewed here[96]) and have been used to identify a number of GPCR structures[9, 19, 20, 22-24, 26-35]. However, their excessive size and the limited mobility of the incorporated membrane proteins prevents them from being used in solution NMR spectroscopy so far.

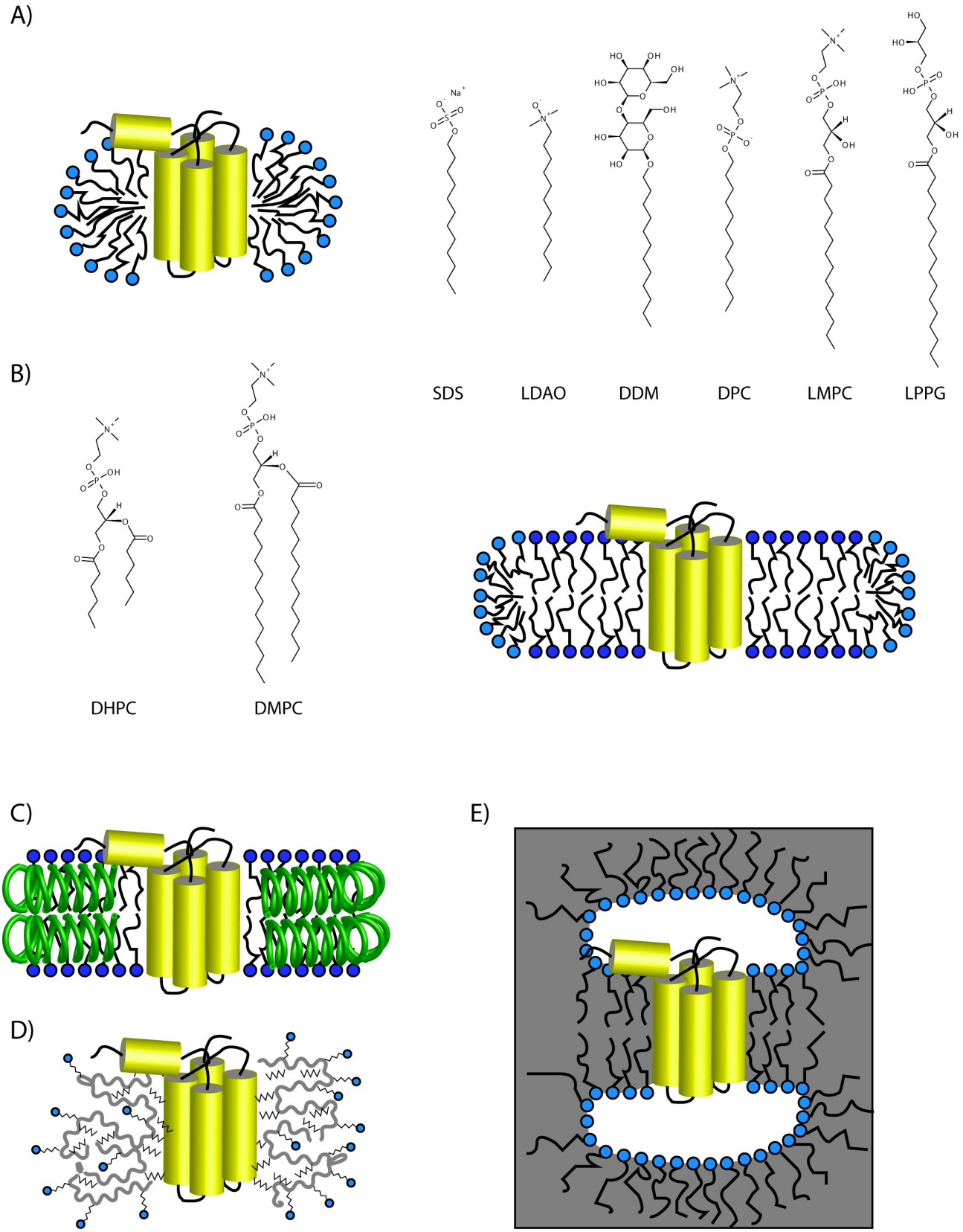


Figure 1.1: Membrane mimetics

Schematic representations of membrane mimetics: Micelles with most commonly used detergents (A), bicelles with detergents (B), nanodiscs (C), amphipols (D) and reverse micelles (E) are shown.

1.1.3 The flexibility problem

As mentioned before, GPCRs possess an inherent flexibility, due to the dynamical nature of the receptors and imperfect lipid-protein contacts as well as the mode of their action, which requires large conformational changes, that makes structure determination quite challenging. Successful attempts to solve this problem for crystallographic studies have included thermostabilization of receptors[52, 97, 98], truncations and deletions of flexible regions, or substitutions by small crystallizable proteins. Especially the substitution of ICL3 with the T4 lysozyme proved to be quite rewarding[16, 22, 28, 29, 32, 99], even though the replacement of ICL3, which plays a vital role in signal proliferation, renders the fused receptor inactive.

On the other hand, the flexibility of GPCRs presents one of the most interesting aspects to investigate as it is directly related to their function. Unlike 'snapshot' techniques like crystallography, NMR spectroscopy has the ability to observe dynamic processes. However, dynamic systems can also present additional problems for NMR spectroscopy in the form of exchange-broadening, depending on the time-scale of the dynamic process. Conformational exchange in the nano- to millisecond time range, which overlaps with the NMR time-scale, can lead to severe line-broadening and consequent loss of signals. While there have not been any high resolution GPCR structures determined by solution-NMR to date, the structures of the bacterial receptors sensory rhodopsin[100, 101] and proteorhodopsin[102], which exhibit similar structural properties as GPCRS, have been solved recently. Some of the advances in NMR spectroscopy that can help with the investigation of GPCRs are outlined in section 1.2.4.

1.2 NMR spectroscopy

1.2.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful analytical techniques in protein structure determination, that capitalizes on quantum mechanical magnetic properties of certain atomic nuclei[103-105]. In NMR spectroscopy, nuclei are placed in a static magnetic field, where they align to form a magnetic dipole. If electromagnetic radiation, in the form of radio frequency pulses, is applied at the right frequency, it is possible to achieve excitation of specific nuclei, which can be observed at a specific resonance frequency. This resonance frequency is dependent on the strength of the magnetic field as well as the specific properties of the observed nucleus and is therefore routinely expressed as a normalized chemical shift scale in ppm. While

equivalent nuclei in the same field should exhibit the exact same chemical shift, differences in the local environment slightly affect the resonance frequency, which in turn facilitates the deduction of information about the local chemical environment . One of the most compelling aspects of NMR spectroscopy for structure determination is it's temporal resilience, allowing for the detection of unstructured and flexible regions, as well as the characterization of dynamical systems and transient interactions.

Conveniently, the most important nuclei in NMR spectroscopy are ^{1}H, ^{13}C and ^{15}N, which also constitute almost all atoms in proteins. Whereas ^{1}H is the most naturally abundant hydrogen isotope, ^{13}C (1.07% natural abundance) and ^{15}N (0.37%) have to be enriched in proteins to be efficiently observed. Isotopic labeling therefore is a critical requirement for protein NMR.

1.2.2 Correlation Experiments

Via scalar or dipolar couplings it is possible to transfer magnetization of one excited nucleus onto another, effectively correlating the resonances of two or more spins. This creates the opportunity to, on the one hand, connect sequentially or spatially proximal nuclei, and, on the other hand, eliminate spectral overlap by the addition of further frequency dimensions. Magnetization can be transferred between practically all NMR-active nuclei, albeit with quite different efficiencies depending on specific characteristics of the relevant nuclei and the magnitude of the involved couplings. Several transfers can be conducted in sequence, resulting in the possibility to obtain experiments in two, three or even more dimensions, correlating a multitude of nuclei to each other. However, every magnetization transfer step results in loss of signal dependent on T_2 relaxation rate, setting an effective limit to the number of observable dimensions, especially for larger molecules. While 4- and 5-dimensional spectra are feasible for small or unfolded proteins[106-108], the measurement of larger proteins or complexes becomes significantly more challenging due to the loss in signal-to-noise ratio. This loss can be somewhat compromised with a reduction in observed data points and respective resolution, which in turn may not be quite as substantial due to the information gained by the addition of further dimensions.

One of the simplest and at same time most important correlation spectrum is the so-called heteronuclear single-quantum coherence spectrum (HSQC). In [$^{15}N^{1}H$]-HSQC experiments ^{1}H nuclei are correlated to ^{15}N nuclei to yield a two-dimensional spectrum with the chemical shifts of ^{1}H on one and ^{15}N on the other axis. As a result, a cross-peak for each covalently bound ^{1}H-^{15}N bond

can be observed. In protein NMR, [^{15}N,^{1}H]-HSQCs exhibit one peak for every backbone amide bond except proline residues, as well as a peak for each indole N-H of Trp residues and a geminal pair of Asn and Gln sidechain amides. HSQC spectra exhibit distinctive patterns for each protein ('fingerprint') and are invaluable as a quick quality control tool, even without assignment. Whereas, signal dispersion provides information about the folding of a protein (well folded proteins exhibit signals from 6-10ppm, while unstructured proteins show very little dispersion between 8-8.5ppm for backbone amides), line-widths report on possible aggregation or conformational exchange (broad lines indicate larger proteins and therefore aggregation). Furthermore, it is also possible to record proton-carbon correlations in the form of [^{13}C,^{1}H]-HSQC experiments, but as they express much smaller signal dispersion, especially in the Cα and Cβ regions, and are therefore not as characteristic as their ^{15}N counterparts, they usually need to be assigned to obtain helpful information. While ^{13}C,^{1}H-correlation experiments such as methyl-TROSY (*vide infra*) can be used to obtain valuable information, they usually require more complicated and expensive labeling techniques, such as reverse methyl labeling (see Section 1.2.4.1).

One major breakthrough in correlation spectroscopy has been the development of Transverse Relaxation Optimized Spectroscopy (TROSY)[109, 110]. The line-widths of NMR peaks are defined by transverse relaxation ($\Delta \nu_{1/2} = 1/(\pi T_2)$). As particles get larger, their correlation time increases, leading to broader and broader lines, effectively demonstrating the size limitation for NMR spectroscopy. The most common pathways for ^{15}N-^{1}H relaxation at high fields are dipole-dipole interactions and chemical shift anisotropy. In experiments where H-N doublets are observed, such as in non-decoupled H-N HSQC spectra, the descriptive terms include additive as well as subtractive elements for dipole-dipole interactions and chemical shift anisotropy (CSA), which at high field strengths (where the CSA effect is strong enough) results in one severely broadened (additive) and one still relatively sharp (subtractive) peak. In TROSY experiments, the broadened components are filtered out and only the optimal component is preserved, leading to significantly improved line-widths compared to regular HSQC spectra.

While the development of TROSY spectroscopy represents an enormous aid in regards of the size limitation of solution NMR spectroscopy, several prerequisites must be fulfilled to obtain the optimal TROSY effect. First of all, the TROSY effect is dependent on strong influence from chemical shift anisotropy, which effectively limits its scope to nuclei with large enough CSA, such as nitrogen atoms. For ^{15}N,^{1}H-correlation spectra, the most favorable field strength for CSA is between 900 and 1100 MHz, which means that the TROSY effect observed on a smaller, for

example 700MHz spectrometer might not be pronounced enough to offer a considerable improvement. Furthermore, TROSY does not cancel out long-range dipolar coupling of protons, which often demands complete perdeuteration of the protein. The development of TROSY nevertheless signifies a major technological advancement, especially for the elucidation of large proteins or complexes. In addition to using TROSY experiments as a substitute for standard [^{15}N,^{1}H]-HSQC spectra, TROSY-based versions of a multitude of multidimensional experiments are widely available, usually exploiting TROSY optimization for the H,N transfer steps. Even though CSA of C-H groups is not effective enough to create a noteworthy TROSY effect, other relaxation interference effects have been identified for methyl and methylene protons[111-113], resulting in the development of resolution optimized methyl-TROSY experiments, which represent great help in terms of sidechain assignment.

1.2.3 Assignment

To specifically connect HSQC peaks to their corresponding residue in proteins, it becomes necessary to obtain more correlations between the amide and the other nuclei in the respective residues and to subsequently link individual amino acids together. For backbone assignments, resonances of amide bonds (NH) need to be correlated to their respective carbonyl (CO), Cα (CA) and Cβ (CB) resonances, as well as their sequential neighbors[114, 115]. Therefore, various three-dimensional triple-resonance spectra are recorded in pairs, one experiment usually correlating the amide moiety with the desired carbon shifts of the preceding residue (*I-1*) and a complementary experiment containing the information of the preceding residue as well as it's own[116]. For example, HNCO experiments report on the preceding carbonyl shift (CO-1), while HN(CA)CO experiments provide information about the preceding as well as the current carbonyl shift (CO and CO-1). Analogous experiments can be recorded for Cα and Cβ resonances in the form of HN(CO)CACB and HNCACB spectra. A schematic representation of the most common backbone experiments can be found in Figure 1.2. By matching the obtained resonances, it is possible to propagate linking neighboring residues and effectively identify long stretches of sequential amino acid residues. As most amino acids have characteristic chemical shifts (Ser, Thr, Gly, Ala have very distinct shifts, whereas the shifts of comparable amino acid, such as Ile and Leu or Asn and Asp are quite similar), it is now possible to successfully map each NMR signal to the corresponding residue in a protein sequence. As prolines only posses tertiary amides, however, they can cause gaps in the assignment, which for most cases does not present a big problem, as the diversity in assigned fragments is usually large enough to lead to unambiguous assignment.

It has to be mentioned again, that the practicability of three-dimensional NMR spectra is highly dependent on protein size, number of magnetization transfer steps and the efficiency of relaxation pathways of the nuclei involved in recording of the spectra. As a rule of thumb, it can be deferred that short three-dimensional experiments that observe only one resonance in the third dimension work better on faster relaxing (i.e. larger) proteins than experiments that include more third dimension resonances or magnetization transfer pathways, e.g. the HNCA is much more sensitive than the HNCACB. Relaxation of carbon nuclei that have protons attached, such as Cα, relax considerably faster than carbonyl nuclei, at least at 700 MHz, and therefore experiments that correlate backbone nuclei (of the predecessor) to the amide resonances via CO usually present better spectra than experiments that involve Cα. While HNCO spectra usually exhibit relatively high signal-to-noise ratios, even for large proteins with long correlation times corresponding up to almost 100kDa, HN(CA)CO spectra already display significantly less signal. The selection for additional peaks in the third dimension, for example when comparing the feasibility of HNCA versus HNCACB experiments, further reduces signal intensities, leading to a limited set of experiments that are still practicable for large proteins or protein complexes. Large membrane proteins, for example, at best only allow for HNCO/HN(CA)CO and HNCA/HN(CO)CA experiments, which can lead to significant difficulties in assignment, as it can be quite challenging to unambiguously distinguish residues based on only a limited set of resonances, especially considering the high prevalence of certain hydrophobic amino acids.

An alternative strategy for backbone assignment is the implementation of amino acid-type selective experiments[117]. With this method, several ^{15}N, ^{1}H correlation spectra similar to [^{15}N,^{1}H]-HSQCs, that specifically select for only certain amino acids, are recorded. Amino acid selective spectra are based on H(CCCO)NH experiments, which typically start on nuclei of the sidechain and are then relayed to the corresponding amide resonance. Different amino acids are selected based on the multiplicity certain sidechain groups (CH_2 versus CH_3) and chemical shifts as well as topology of the sidechain. Similarly to 'classical' backbone assignment, a set of two different experiments is recorded for each amino acid, one containing information of the succeeding residue called an (*i+1*)-HSQC and one additionally including information about the detected amino acid, (*i, i+1*)-HSQC, thus sequentially linking two residues. As it is not possible to select for each amino acid, due to the high similarity of some sidechains, several sets of selective labeling experiments often become necessary to unambiguously distinguish and assign backbone resonances. Amino acid type selective experiments therefore present a viable backbone assignment strategy for most soluble proteins,

provided they possess a relatively standard amino acid distribution. Their usefulnes for larger proteins or membrane proteins with limited amino acid distributions is less obvious.

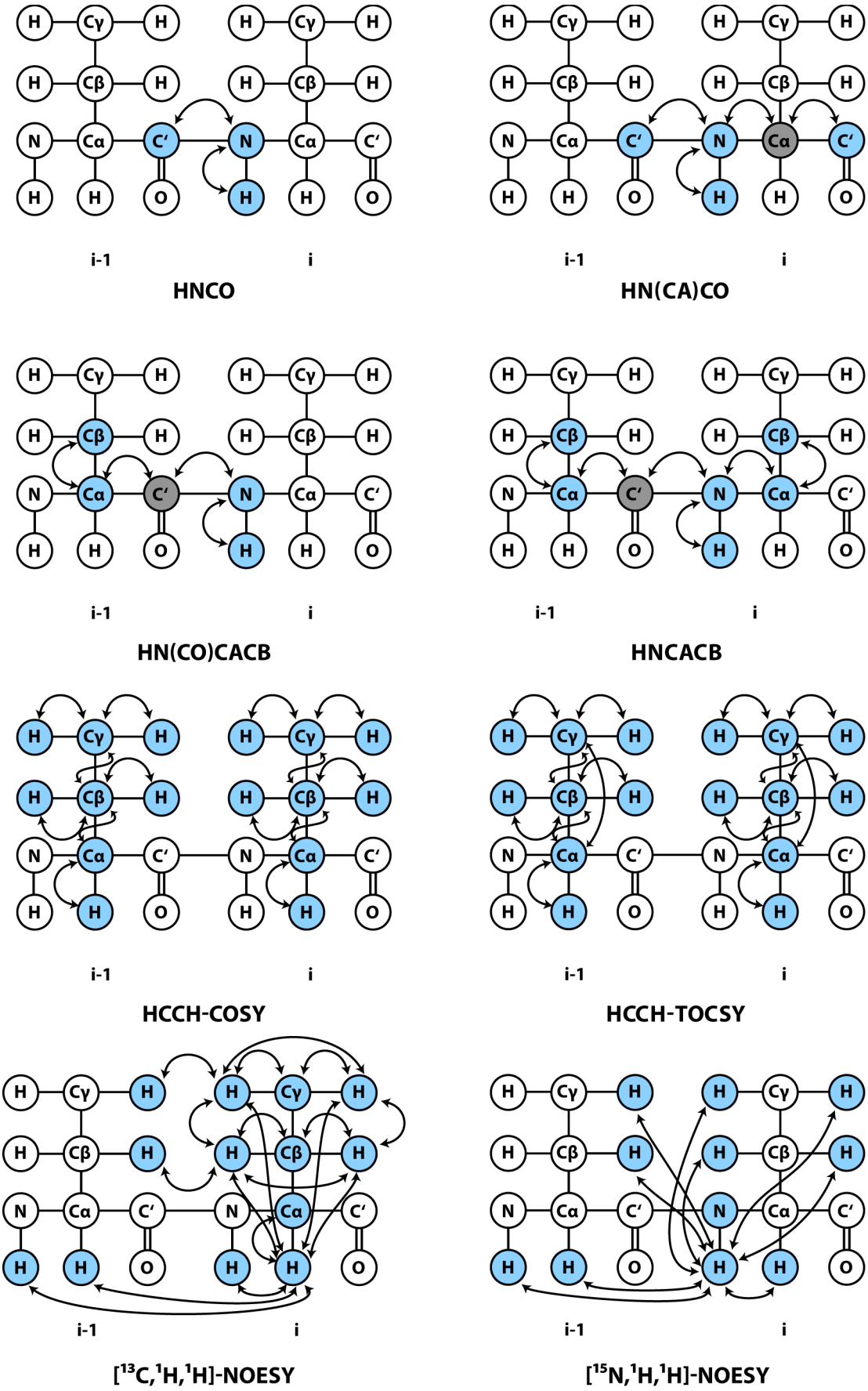


Figure 1.2: Pathways of NMR experiments

Schematic representation of the most common experiments for NMR assignment. Observable nuclei for each experiment are shown in blue; nuclei that are involved in magnetization transfer, but are not shift labeled are indicated in grey

Whereas secondary structure can be predicted solely on chemical shifts of the backbone assignment[118], it is indispensable to also assign the resonances of sidechain atoms, especially protons, to calculate protein structures. Sidechain resonance assignment is possible using the ^{15}N-edited spectra H(CCCO)NH and hC(CCO)NH or ^{15}N-TOCSY experiments, which directly connect the proton and carbon sidechain resonances of the preceding or native residue, respectively, to the observed amide peak. In combination with ^{13}C-edited experiments such as HCCH-COSY, which permits stepwise discrimination of single sidechain atoms, unambiguous assignment is feasible. The correlation of sidechain resonances to amide protons, however, involves a high number of magnetization transfer steps with additive signal loss for each transfer. Therefore, ^{15}N-edited experiments require fast tumbling and slow relaxation rates, effectively limiting their use to small soluble proteins. Consequently, mainly ^{13}C-edited (based on the [^{13}C,^{1}H]-HSQC), three-dimensional triple resonance experiments, such as (H)CCH-COSY and (H)CCH-TOCSY are utilized for the assignment of membrane proteins[119, 120]. (H)CCH-COSY and (H)CCH-TOCSY experiments permit the assignment of carbon resonances throughout the entire spin system (I.e. amino acid), while H(C)CH-COSY and H(C)CH-TOCSY describe the proton analogues. Additionally, ^{15}N-resolved [^{1}H,^{1}H]-NOESY and ^{13}C-resolved [^{1}H,^{1}H] -NOESY spectra provide spatial information of proton nuclei. Nuclear Overhauser effect (NOE) resonances therefore provide the primary tool for protein structure determination. Schematic representations of some experiments can be seen in Figure 1.2.

1.2.4 Membrane Protein NMR

Membrane protein NMR presents several additional obstacles compared to NMR spectroscopy of soluble proteins. First of all, particle size is a limiting factor in NMR spectroscopy. As stated above, membrane proteins require to be stabilized by the presence of a membrane mimetic. The resulting protein-membrane mimetic complex significantly increases the particle size, potentially up to a multiple of the size of the original protein. This increased size results in an increase in correlation time and subsequent increased R_2 rates resulting in line-broadening. It is therefore imperative to identify a suitable membrane mimetic, that sufficiently stabilizes the protein while at the same time maintains a manageable particle size. Furthermore, the internal protein dynamics of membrane proteins present severe complications. Conformational changes set in the nano- to millisecond time-range result in exchange-broadening, frequently rendering portions of the protein undetectable. This problem is especially exaggerated for intrinsically flexible systems such as GPCRs. Moreover, α-helical integral membrane proteins demonstrate relatively little signal dispersion, resulting in overlapping peaks. As the termini of membrane proteins are often

unstructured, they result in comparatively strong signals, which, in case of peak overlap, can easily conceal crucial peaks from comparably weak helical signals. The transmembrane regions naturally also favor the occurrence of hydrophobic amino acids, such as alanine, isoleucine and leucine, further complicating unambiguous resonance assignment. Additionally, even if the membrane protein can be assigned successfully, structural determination remains a challenge, as the predominant source for structure elucidation, NOEs, often can not be observed due to the above mentioned flexibility and line-broadening effects. In addition, NOEs might be present but could stem from unassigned sidechain resonances.

In conclusion, structural characterization of membrane proteins using conventional solution NMR spectroscopy is often not feasible. However, recent advancements in the development of NMR methods using sparse, but more relevant data, which can still be recorded and evaluated, as well as improved computational tools, still provide a viable way to gain structural information. As mentioned above, sidechain assignment of membrane proteins is quite challenging and time consuming, rarely providing a sufficient number of long-range contacts, which are necessary for high-resolution structure determination. While recent advances in the development of *ab initio* structure calculation tools such as ROSETTA[121, 122] have greatly aided membrane protein structure determination, it might be necessary to incorporate additional techniques such as small-angle x-ray scattering (SAXS)[123, 124] and cryo-electron microscopy (cryo-EM)[125].

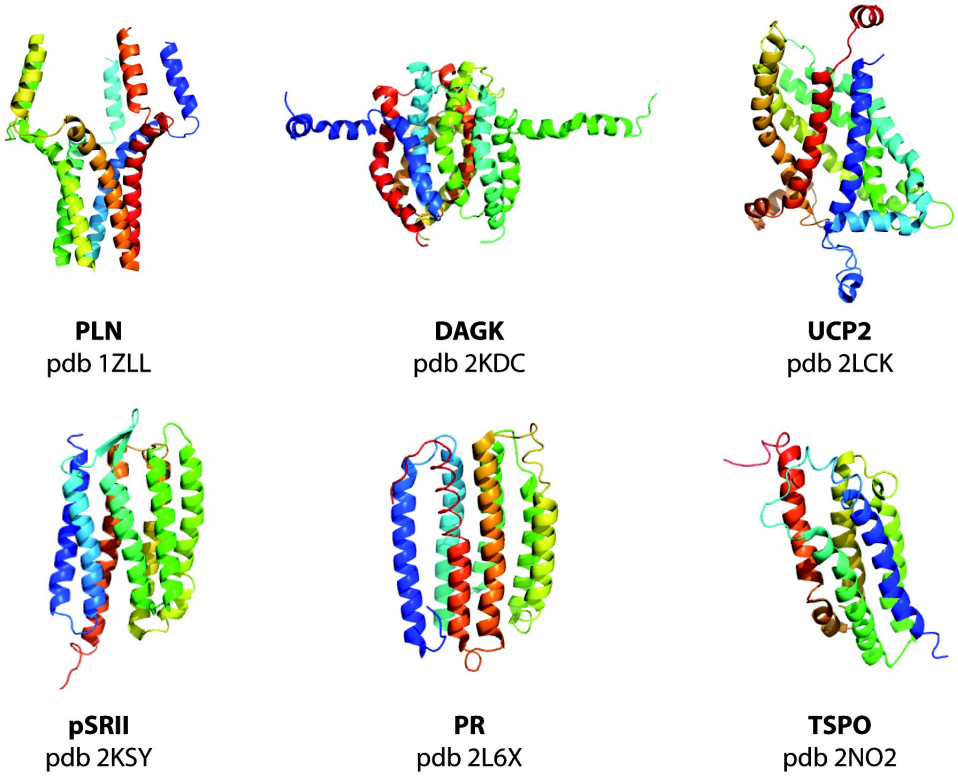


Figure 1.3: Membrane protein structures solved with sparse sampling data

Selection of structures of large membrane proteins using sparse NMR data.

However, it is also possible to determine membrane protein structures purely based on NMR data. A combination of structure prediction tools such as ROSETTA[126] with sparse NMR data, such as paramagnetic relaxation enhancement (PRE)[127, 128], residual dipolar coupling (RDC)[129-131], pseudo-contact shifts (PSC)[122, 132, 133] and selective NOEs[121, 134-136] has proven to be a viable strategy for membrane protein structure determination[126]. Suitable sparse NMR data may still provide information of long-range contacts or global orientation without the need for complete sidechain assignment. Using sparse NMR data, it has been possible to determine the structures of several large α-helical proteins, containing multiple transmembrane helices and domains, such as the 30kDa homopentamer human phospholamban (PLN)[137], the 43kDa homotrimer diacylglycerol kinase (DAGK)[138], the 33kDa mitochondrial uncoupling protein 2 (UCP2)[139], the 18kDa mitochondrial translocator protein (TSPO)[140] as well as the seven transmembrane bacterial receptors sensory rhodopsin II (pSRII)[100] and proteorhodopsin (PR)[102], all of which can be found in Figure 1.3. The following sections will describe the experimental details as well as advantages and limitations of methods that are employed in membrane protein NMR.

1.2.4.1 Isotope Labeling

As mentioned before, it is indispensable to uniformly substitute the nitrogen and carbon atoms with the respective ^{15}N and ^{13}C isotopes in order to record two-, three- and more-dimensional NMR spectra. ^{1}H already is the most abundant hydrogen isotope and therefore does not need to be supplanted. However, protons also have the biggest effect on transverse relaxation rates and hence lead to loss of signal. By substituting proton with deuterium isotopes, dipole-dipole interactions, which are the primary agent for transverse relaxation, are reduced, therefore lowering relaxation rates significantly and leading to substantial increases in signal-to-noise ratios. Protein expression in *E. coli* expression systems, for example, can be carried out in D_2O, which normally results in a perdeuteration rate of at least 80% (when using non-deuterated glucose), with mostly methyl groups or other sidechain groups located further away from the backbone still being protonated. Under standard measuring conditions, amide protons rapidly back-exchange to ^{1}H, which makes the application of [^{15}N,^{1}H]-HSQC or [^{15}N,^{1}H]-TROSY based spectra feasible. This is especially useful for HNCACB and HN(CO)CACB experiments, as the transverse relaxation times for Cα and Cβ nuclei is increased by about an order of magnitude. As a potential disadvantage, complete perdeuteration also eliminates the potential of obtaining experiments that involve or even start on aliphatic proton atoms, like sidechain or NOESY experiments. It is also possible to express proteins with random fractional deuteration, e.g. 50% deuteration, which on the one hand improves relaxation rates, while preserving a sufficient density of protons for detecting NOEs. It has to be mentioned though, that the prolongation of relaxation rates in fractionally deuterated proteins in some cases is not significant enough to result in decidedly improved spectra. Furthermore, the presence of different isotopomers in partially deuterated proteins result in heterogenous peak shapes further complicating the determination of peak positions.

A much superior, while also more complicated strategy to improve sample characteristics is the use of selective isotope labeling. In segmental isotope labeling, only certain parts of a protein are visible via NMR spectroscopy. Therefore, two or more segments need to be expressed individually under different labeling conditions and subsequently fused together, either chemically[141, 142], enzymatically[143, 144] or transiently through electrostatic interactions[145]. Another approach to utilize selective isotope labeling can be achieved by only labeling specific amino acids. While enzymatic scrambling during protein expression does not allow the isotopic substitution of any amino acid, certain amino acids are quite amendable to this technique. The most common application is the selective reverse methyl labeling of isoleucine, leucine and valine (as well as

alanine, methionine and threonine to a lesser extent). With this approach, only one methyl group in each Ile, Leu and Val residue is protonated, while the remaining protein remains perdeuterated. Even though unambiguous assignment of the methyl resonances might still prove quite challenging, especially for membrane proteins that naturally present a high percentage of the hydrophobic residues Ile, Leu and Val, this technique enables the observation of tertiary structure interactions in the form of methyl-methyl and methyl-amide NOEs, which is particularly important for large proteins. For this technique, it is also imperative to achieve complete background deuteration, so in addition to D_2O, also a perdeuterated carbon source (usually $^{13}C,^{2}H$-glucose or $^{13}C,^{2}H$-glycerol) has to be supplemented during protein expression. Furthermore, when working with membrane proteins, it may also be necessary to use perdeuterated detergents, which presently pose limitations in terms of price and availability.

The implementation of cell-free expression also expands the scope of labeling schemes as it effectively limits isotope scrambling. Therefore, techniques such as stereo-array isotope labeling (SAIL)[146], where one of the distereotopic methylene protons is stereospecifically replaced by deuterium (in addition also equivalent ring protons or equivalent Me groups) , are feasible. However, the price of individual SAIL amino acids often are prohibitive.

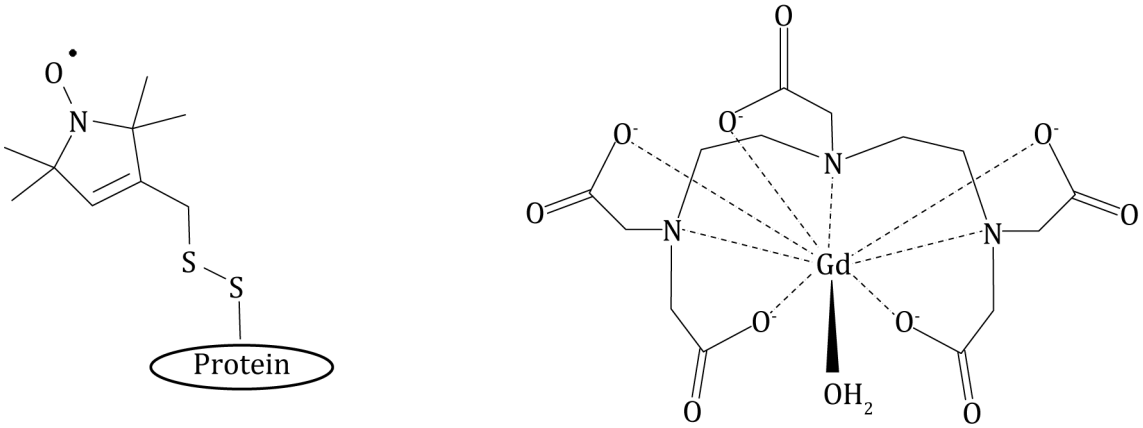
1.2.4.2 PRE and PSC

As was described above, the acquisition of long range distance restraints for the structure calculation of membrane proteins can be quite challenging. Considering the low number of long-range NOEs, paramagnetic relaxation enhancement (PRE) experiments present a very useful tool to gather long distance information[128]. Analogous to NOE experiments, paramagnetic relaxation is a very effective relaxation pathway that attenuates signals from spins in a distance-dependent manner, however it affects residues at a significantly farther distance (up to 30 Å) than NOEs (up to 6Å) and can therefore tremendously augment long distance information. Furthermore, signal attenuation caused by paramagnetic relaxation can be detected considerably easier than the intrinsically weak NOEs from long-range interactions. Paramagnetic spinlabels contain unpaired electrons, usually in the form of stabilized nitroxide groups or paramagnetic metal ions, that can be attached to the proteins via a linker, and that increase the relaxation rates of close nuclei. The data acquired in PRE experiments can thus directly be converted in structure calculation restraints[102, 128].

PRE measurements, of course, demand a feasible strategy to introduce paramagnetic spinlabels at unique positions. Most commonly, native cysteine residues are substituted to equivalent amino acids such as serine, with the subsequent reintroduction of single cysteines to which the spinlabel, *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) as the most common, can be attached. However, the elimination of native cysteine residues can interfere with protein folding and consequently activity, which means that the activity of the mutated proteins should be confirmed.

Another strategy to gather topological information through paramagnetic relaxation is the addition of a soluble, inert paramagnetic probe, such as Gd-(DTPA-BMA) (Gadolinium diethylenetriamine-pentaacetic acid bis-methyl amine) to the NMR sample. As it results in attenuation of water accessible residues, this technique can be used to identify protein binding and dimerization interfaces in soluble proteins and proves to be especially valuable for the characterization of membrane protein insertion into micelles or other membrane mimetics[147].

Additionally to paramagnetic probes that result in fast relaxation, some agents do not cause signal attenuation, but perturb NMR chemical shifts through the induction of pseudo-contact shifts (PCS). For this purpose, analogous to PRE experiments chelated Co^{2+} or lanthanide(III) ions are attached to certain sites in the protein, which alter the chemical shifts of residues that are close in space (up to 40Å) in a distance- and orientation-dependent manner. It has been shown that PSC can greatly aid in structure determination of proteins[148].



MTSL Gd-(DPTA-BMA)

Figure 1.4: Spinlabel structures

Chemical structures of the paramagnetic spinlabels MTSL (left) and Gd-(DPTA-BMA)

1.2.4.3 RDC

An additional method to collect structural information is to record Residual Dipolar Coupling (RDC)[149, 150]. RDCs provide data about the orientation of a vector connecting two nuclei, relative to the overall orientation of the protein in an alignment medium, leading to global structure information. In solution-NMR, proteins tumble freely in solution, which causes dipolar interactions to be averaged out isotropically. However, it is possible to manipulate the system so that partial alignment of proteins can be achieved, which in turn results in formation of RDCs. Paramagnetic metal ions will be affected by the strong magnetic field in an NMR spectrometer and can be attached to the protein, to achieve marginal alignment. The more common strategy, however, is the addition of macromolecular alignment media, such as bicelles, phage particles or lamellar media consisting of ether/alcohol mixtures, to the NMR sample[151]. These media will assume a preferred orientation relative to the magnetic field, thereby impeding the tumbling of not perfectly spherical proteins.

However, the analysis of membrane proteins presents some additional challenges, as membrane mimetics tend to interact with the alignment medium, consequently resulting in over-alignment or severe line-broadening since the alignment medium presents a very slowly tumbling entity. One of the most effective systems suitable for membrane proteins are G-tetrads[152], which consist of 2'-deoxyguanylyl-(3',5')-2'-deoxyguanosine (dGpG) and which form in the presence of potassium ions. K-dGpG stacks align perpendicular to the magnetic field, due to their high net negative charge and therefore can be employed in combination with negatively charged macromolecules such as detergent micelles.

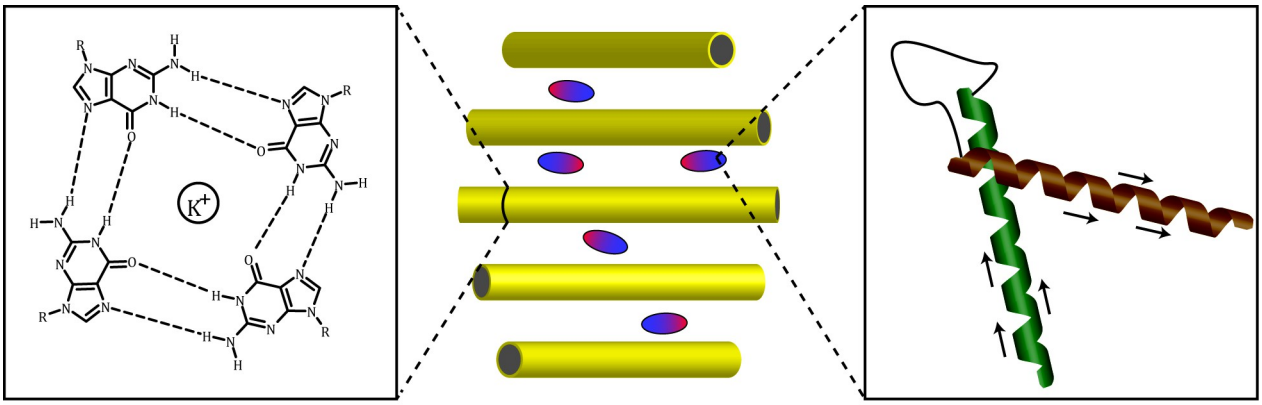


Figure 1.5: RDC scheme

Schematic representation of residual dipolar coupling experiments. Chemical structure of G-tetrade (left), partial protein alignment in presence of macromolecular alignment media (center) and approximate relative orientation of N-H vectors (right) are shown.

1.3 Scope and Project Goals

As described above, structural investigation of GPCRs is a highly interesting topic. As the structure determination of whole GPCRs using NMR spectroscopy is still not yet feasible, we have utilized a 'divide-and-conquer' approach to split GPCRs in fragments of one or more transmembrane helices (TM), which presents several interesting aspects in the study of GPCRs. In contrast to most soluble proteins, membrane proteins exhibit secondary and possibly also tertiary strucutre that make the study of protein fragments worth to pursuit. Whereas soluble proteins derive their structure in general from the formation of a hydrophobic core, which would not fold properly if parts of a domain are missing, membrane proteins usually do not possess a hydrophobic core and are rather stabilized by the membrane environment, a hydrophobic exterior, that in concert with the amino acid sequence defines the secondary structure. The two-stage membrane protein folding mechanism by Popot and Engelmann[153] proposes the initial formation of α-helices upon partitioning into the membrane, which diffuse within the bilayer and form native tertiary contacts upon addition of the remaining helices. The study of receptor fragments or more accurately the addition of further helices to shorter fragments, enables the detailed investigation and characterization of stabilizing tertiary contacts between helices (illustrated in Figure 1.6), which in turn can lead to valuable insight about receptor folding and biosynthesis.

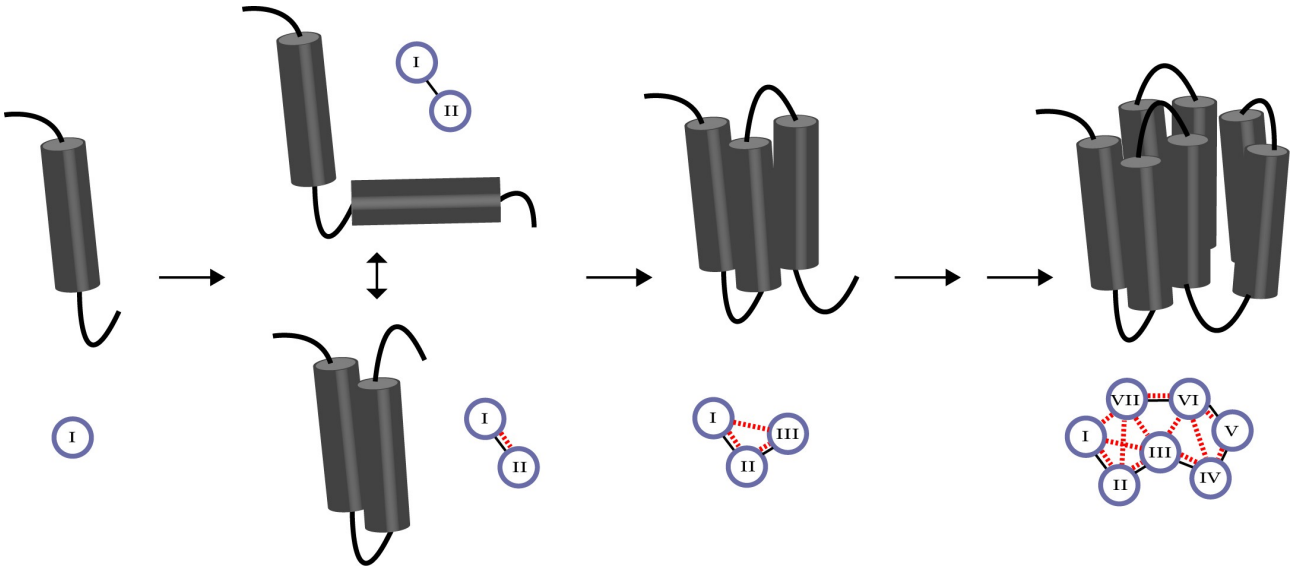


Figure 1.6: Study of interhelical contacts

Schematic representation of the study of interhelical contacts between transmembrane helices. The formation of interhelical contacts is indicated by dotted red lines.

The utilization of solution-NMR spectroscopy additionally permits the observation of dynamic processes, which can be used to define flexible regions and consequently gather information about regions, which are important for receptor folding and possibly function. As chemical shifts are mostly dependent on the local environment, the assignments of shorter fragments can be transferred to longer ones, thereby greatly facilitating the assignment of large fragments and ultimately hopefully whole receptors, as is illustrated in Figure 1.7. Another aspect of interest would be the investigation of reconstituted split receptors, where it would be possible to segmentally label different parts of the receptor to significantly reduce effective relaxation rates and simplify spectra. Preliminary studies in the guided reconstitution of receptors can be found in the work by Cohen *et al*[154].

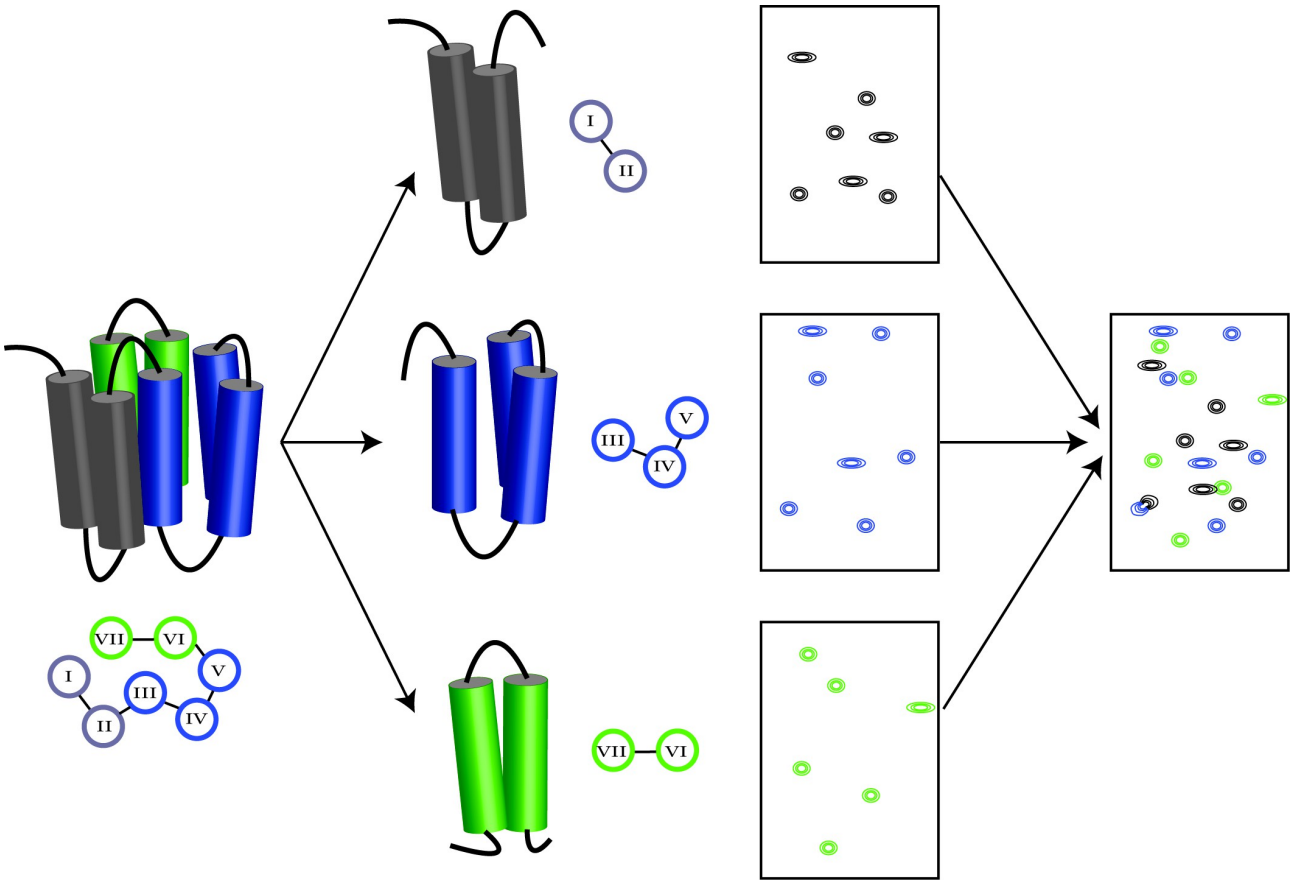


Figure 1.7: Assignment transfer scheme

Schematic representation of NMR assignment transfer from GPCR fragments.

The following chapters will describe my work on fragments from the yeast GPCR Ste2p. A snake-plot representation of Ste2p can be found in Figure 1.8. Ste2p is a fungal mating pheromone receptor (Class D GPCR) and initiates mating in yeast upon binding the pheromone α-factor[155]. To

study structural preferences, interhelical contacts, as well as early receptor folding events, we have designed an approach to systematically increase the number of transmembrane helices in receptor fragments. The fragments TM1, TM12, TM123 and TM127 containing one, two, and three transmembrane-helices respectively that in the native receptor form contacts were designed, expressed, purified and investigated using a range of biophysical experiments, most notably NMR spectroscopy. To achieve this, different expression designs and strategies were implemented. Expression of the desired protein initially in *E.coli* was accomplished with a N-terminal fusion sequence, which included the optimization of different heterologous expression strains and conditions as well as the chemical cleavage and subsequent RP-HPLC purification of the fragments. The implementation of different NMR methods, such as the measurement of PREs, required direct expression of the fragments, which led to further expression and purification optimization in order to preserve integrity of the spin label. Samples were studied using NMR spectroscopy by dissolving them in organic solvent/water mixtures or solubilizing them in micelles, using a mixture of LPPG and DPC as the detergents.

Almost complete backbone and partial sidechain assignment could be achieved for all fragments using the NMR experiments described above. Assignment was aided tremendously by the transfer of chemical shifts from shorter to longer fragments, a process which we were able to automate utilizing automated peak picking and NOESY strip matching. Analysis of backbone chemical shifts using the secondary prediction tool TALOSn indicated the correct formation of α-helices at the positions predicted by homology modeling, confirming the validity of the proposed two-stage model of membrane protein folding. The formation of stabilizing tertiary contacts could be observed by expanding the TM1 fragment to TM12, whereas the addition of another helix resulted in unsatisfied interhelical contacts that consequently led to destabilization in TM123 or the presence of multiple conformations in case of TM127. Due to the inherently limited amount of observable long-range interactions in membrane proteins, a series of NMR experiments, including paramagnetic relaxation enhancements, residual dipolar couplings and the use of soluble paramagnetic spinlabels was utilized.

The following three chapters present the results obtained during this PhD project. Chapter 2 will describe the expression, purification and biophysical analysis of Ste2p fragments containing the first three transmembrane helices (TM123). It will include different cloning and expression strategies, as well as NMR spectroscopy data in trifluoroethanol/water mixtures. Chapter 3 will detail the systematic elongation of fragments from one to three transmembrane helices and contrast

structural preferences between the different fragments. Constituting the largest part of work undertaken during this thesis, it will also detail structure determination by NMR spectroscopy and the mapping of structural preferences using sparse NMR data. Chapter 4 will review methods employed by our group to study topological and structural preferences of GPCR fragments as well as early receptor folding. While the results presented in Chapter 2 have been published in their entirety, Chapters 3 and 4 are part of ongoing research and have not been published yet.

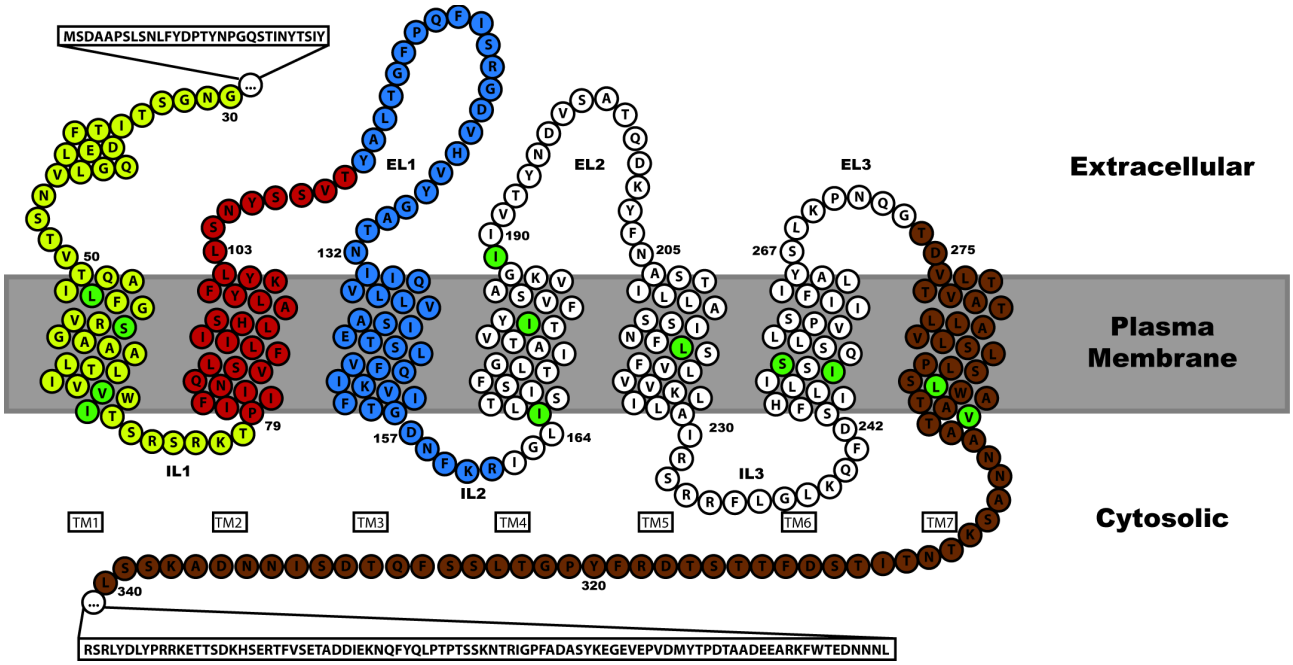


Figure 1.8: Ste2p snake-plot

Snake plot representation of the Ste2p receptor. Residues studied as part of fragments are colored in yellow (TM1), red (TM2), blue (TM3) and brown (TM7). Cysteines and methionines have been replaced by similar residues and are indicated in green.

1.4 References

1. Bockaert J, Pin JP. *Molecular tinkering of G protein-coupled receptors: an evolutionary success.* Embo J. 1999;**18:**1723-9.
2. Pierce KL, Premont RT, Lefkowitz RJ. *Seven-transmembrane receptors.* Nat Rev Mol Cell Biol. 2002;**3:**639-50.
3. Roth BL. *Receptor systems: will mining the receptorome yield novel targets for pharmacotherapy?* Pharmacology & therapeutics. 2005;**108:**59-64.
4. Wise A, Gearing K, Rees S. *Target validation of G-protein coupled receptors.* Drug Discov Today. 2002;**7:**235-46.
5. Bockaert J. *[G-protein coupled receptors. Nobel Prize 2012 for chemistry to Robert J. Lefkowitz and Brian Kobilka].* Medecine sciences : M/S. 2012;**28:**1133-7.
6. Bockaert J. *[G-protein-coupled receptors: general features and activation mechanisms].* Bulletin de l'Academie nationale de medecine. 2012;**196:**1765-75.
7. Russ WP, Engelman DM. *The GxxxG motif: a framework for transmembrane helix-helix association.* J Mol Biol. 2000;**296:**911-9.
8. Hofmann K, Scheerer P, Hildebrand P, Choe H, Park J, Heck M, et al. *A G protein-coupled receptor at work: the rhodopsin model.* Trends Biochem Sci. 2009**:** .
9. Scheerer P, Park J, Hildebrand P, Kim Y, Krauss N, Choe H, et al. *Crystal structure of opsin in its G-protein-interacting conformation.* Nature. 2008;**455:**497-502.
10. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, Ernst OP. *Role of the conserved NPxxY(x)(5,6)F motif in the rhodopsin ground state and during activation.* Proceedings of the National Academy of Sciences of the United States of America. 2003;**100:**2290-5.
11. van der Kant R, Vriend G. *Alpha-Bulges in G Protein-Coupled Receptors.* International Journal of Molecular Sciences. 2014;**15:**7841-64.
12. Kofuku Y, Yoshiura C, Ueda T, Terasawa H, Hirai T, Tominaga S, et al. *Structural basis of the interaction between chemokine stromal cell-derived factor-1/CXCL12 and its G-protein-coupled receptor CXCR4.* J Biol Chem. 2009;**284:**35240-50.
13. Foord SM. *Receptor classification: post genome.* Curr Opin Pharmacol. 2002;**2:**561-6.
14. Katritch V, Cherezov V, Stevens RC. *Diversity and modularity of G protein-coupled receptor structures.* Trends Pharmacol Sci. 2012;**33:**17-27.
15. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science. 2000;**289:**739-45.
16. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, et al. *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science. 2007;**318:**1258-65.
17. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, et al. *The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.* Science. 2008;**322:**1211-7.
18. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP. *Crystal structure of the ligand-free G-protein-coupled receptor opsin.* Nature. 2008;**454:**183-7.

19. Warne T, Serrano-Vega M, Baker J, Moukhametzianov R, Edwards P, Henderson R, et al. *Structure of a beta1-adrenergic G-protein-coupled receptor.* Nature. 2008;**454:**486-91.

20. Chien EYT, Liu W, Katritch V, Abagyan R, Kuhn P, Cherezov V, et al. *Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists.* Science. 2010;**330:**1066-71.

21. Chien EYT, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, et al. *Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist.* Science. 2010;**330:**1091-5.

22. Zhang C, Srinivasan Y, Arlow D, Fung J, Palmer D, Zheng Y, et al. *High-resolution crystal structure of human protease-activated receptor 1.* Nature. 2012;**492:**387-92.

23. Wu B, Chien E, Mol C, Fenalti G, Liu W, Katritch V, et al. *Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists.* Science. 2010;**330:**1066-71.

24. Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, et al. *Structure of the Adenosine A(2A) Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine.* Structure. 2011;**19:**1283-93.

25. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, et al. *Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation.* Nature. 2011;**474:**521-5.

26. Rasmussen S, Choi H, Fung J, Pardon E, Casarosa P, Chae P, et al. *Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor.* Nature. 2011;**469:**175-80.

27. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, et al. *Structure of the human histamine H1 receptor complex with doxepin.* Nature. 2011;**475:**65-70.

28. Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, et al. *Structure of the δ-opioid receptor bound to naltrindole.* Nature. 2012;**485:**400-4.

29. Haga K, Kruse A, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, et al. *Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist.* Nature. 2012;**482:**547-51.

30. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, et al. *Crystal structure of a lipid G protein-coupled receptor.* Science. 2012;**335:**851-5.

31. Kruse A, Li J, Hu J, Kobilka B, Wess J. *Novel insights into m3 muscarinic acetylcholine receptor physiology and structure.* J Mol Neurosci. 2014;**53:**316-23.

32. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, et al. *Crystal structure of the μ-opioid receptor bound to a morphinan antagonist.* Nature. 2012**:**1-7.

33. Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, et al. *Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic.* Nature. 2012;**485:**395-9.

34. White J, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, et al. *Structure of the agonist-bound neurotensin receptor.* Nature. 2012;**490:**508-13.

35. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, et al. *Structure of the human κ-opioid receptor in complex with JDTic.* Nature. 2012;**485:**327-32.

36. Huang J, Chen S, Zhang JJ, Huang X-Y. *Crystal structure of oligomeric β1-adrenergic G*

protein–coupled receptors in ligand-free basal state. Nat Struct Mol Biol. 2013;**20:**419-25.

37. Hollenstein K, Kean J, Bortolato A, Cheng RKY, Doré AS, Jazayeri A, et al. *Structure of class B GPCR corticotropin-releasing factor receptor 1.* Nature. 2013;**499:**438-43.

38. Siu FY, He M, De Graaf C, Han GW, Yang D, Zhang Z, et al. *Structure of the human glucagon class B G-protein-coupled receptor.* Nature. 2013;**499:**444-9.

39. Wang C, Wu H, Katritch V, Han GW, Huang X-P, Liu W, et al. *Structure of the human smoothened receptor 7TM bound to an antitumor agent.* Nature. 2013;**497:**338-43.

40. Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, et al. *Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain.* Nature. 2014;**511:**557-62.

41. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, et al. *Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator.* Science. 2014;**344:**58-64.

42. Zhang J, Zhang K, Gao Z, Paoletta S, Zhang D, Han G, et al. *Agonist-bound structure of the human P2Y12 receptor.* Nature. 2014;**509:**119-22.

43. Zhang K, Zhang J, Gao Z, Zhang D, Zhu L, Han G, et al. *Structure of the human P2Y12 receptor in complex with an antithrombotic drug.* Nature. 2014;**509:**115-8.

44. Standfuss J, Edwards P, D'Antona A, Fransen M, Xie G, Oprian D, et al. *The structural basis of agonist-induced activation in constitutively active rhodopsin.* Nature. 2011;**471:**656-60.

45. Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD, Schertler GF. *Crystal structure of a thermally stable rhodopsin mutant.* J Mol Biol. 2007;**372:**1179-88.

46. Andrell J, Tate CG. *Overexpression of membrane proteins in mammalian cells for structural studies.* Molecular membrane biology. 2013;**30:**52-63.

47. Wu H, Wang C, Gregory K, Han G, Cho H, Xia Y, et al. *Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator.* Science. 2014;**344:**58-64.

48. Fenalti G, Giguere PM, Katritch V, Huang X-P, Thompson AA, Cherezov V, et al. *Molecular control of δ-opioid receptor signalling.* Nature. 2014;**506:**191-6.

49. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, et al. *Structure of the human histamine H1 receptor complex with doxepin.* Nature. 2011**:**1-8.

50. Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, et al. *G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody.* Nature. 2012.

51. Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, et al. *Structure of the chemokine receptor CXCR1 in phospholipid bilayers.* Nature. 2012;**491:**779-83.

52. Egloff P, Hillenbrand M, Klenk C, Batyuk A, Heine P, Balada S, et al. *Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli.* Proc Natl Acad Sci U S A. 2014;**111:**E655-62.

53. Marullo S, Delavier-Klutchko C, Eshdat Y, Strosberg AD, Emorine L. *Human beta 2-adrenergic receptors expressed in Escherichia coli membranes retain their pharmacological properties.* Proceedings of the National Academy of Sciences of the United States of America. 1988;**85:**7551-5.

54. Weiss HM, Grisshammer R. *Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli.* Eur J Biochem. 2002;**269:**82-92.

55. Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. *Conformational thermostabilization of the 1-adrenergic receptor in a detergent-resistant form.* Proc Natl Acad Sci U S A. 2008**:** .

56. Yeliseev A, Zoubak L, Gawrisch K. *Use of Dual Affinity Tags for Expression and Purification of Functional Peripheral Cannabinoid Receptor.* Protein expression and purification. 2007;**53:**153-63.

57. Chowdhury A, Feng R, Tong Q, Zhang Y, Xie X-Q. *Mistic and TarCF as fusion protein partners for functional expression of the cannabinoid receptor 2 in Escherichia coli.* Protein expression and purification. 2012;**83:**128,Äì34.

58. Kiefer H, Krieger Jr, Olszewski JD, Heijne Gv, Prestwich GD, Breer H. *Expression of an Olfactory Receptor in Escherichiacoli: Purification, Reconstitution, and Ligand Binding ‚Ä†.* Biochemistry. 1996;**35:**16077,Äì84.

59. Cohen L, Fracchiolla K, Becker J, Naider F. *GPCR Structural Characterization: Using Fragments as Building Blocks to Determine a Complete Structure.* Biopolymers. 2014.

60. Haberstock S, Roos C, Hoevels Y, Dötsch V, Schnapp G, Pautsch A, et al. *A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems.* Protein Expression and Purification. 2012;**82:**308-16.

61. Zhou H-X, Cross TA. *Influences of Membrane Mimetic Environments on Membrane Protein Structures.* Annual review of biophysics. 2013;**42:**361-92.

62. Werner K, Richter C, Klein-Seetharaman J, Schwalbe H. *Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy.* J Biomol NMR. 2008;**40:**49-53.

63. Klein-Seetharaman J, Yanamala NV, Javeed F, Reeves PJ, Getmanova EV, Loewen MC, et al. *Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR.* Proc Natl Acad Sci U S A. 2004;**101:**3409-13.

64. Page RC, Moore JD, Nguyen HB, Sharma M, Chase R, Gao FP, et al. *Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins.* J Struct Funct Gen

. 2006;**7:**51-64.

65. Tian C, Karra M, Ellis C, Jacob J, Oxenoid K, Sonnichsen F, et al. *Membrane protein preparation for TROSY NMR screening.* Methods Enzymol. 2005;**394:**321-34.

66. Beel AJ, Mobley CK, Kim HJ, Tian F, Hadziselimovic A, Jap B, et al. *Structural studies of the transmembrane C-terminal domain of the amyloid precursor protein (APP): does APP function as a cholesterol sensor?* Biochemistry. 2008;**47:**9428-46.

67. Oliver RC, Lipfert J, Fox DA, Lo RH, Doniach S, Columbus L. *Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group.* PLoS ONE. 2013;**8:**e62488.

68. Sanders C, S5 onnichsen F. *Solution NMR of membrane proteins: practice and challenges.* Magnetic resonance in chemistry : MRC. 2006;**44 Spec No:**S24-S40.

69. Tian C, Vanoye CG, Kang C, Welch RC, Kim HJ, George AL, et al. *Preparation, Functional Characterization, and NMR Studies of Human KCNE1, a Voltage-Gated Potassium Channel Accessory Subunit Associated with Deafness and Long QT Syndrome(,).* Biochemistry. 2007;**46:**11459-72.

70. Krueger-Koplin RD, Sorgen PL, Krueger-Koplin ST, Rivera-Torres IO, Cahill SM, Hicks DB, et al. *An evaluation of detergents for NMR structural studies of membrane proteins.* J Biomol NMR. 2004;**28:**43-57.

71. Qin X, Liu M, Yang D, Zhang X. *Concentration-dependent aggregation of CHAPS investigated by NMR spectroscopy.* J Phys Chem B. 2010;**114:**3863-8.

72. Sanders CR, Prosser RS. *Bicelles: a model membrane system for all seasons?* Structure. 1998;**6:**1227-34.

73. Vold RR, Prosser RS, Deese AJ. *Isotropic Solutions Of Phospholipid Bicelles - a new membrane mimetic for high-resolution nmr studies of polypeptides.* J Biomol NMR. 1997;**9:**329-35.

74. Fernandez C, Hilty C, Wider G, Guntert P, Wuthrich K. *NMR structure of the integral membrane protein OmpX.* J Mol Biol. 2004;**336:**1211-21.

75. Roosild T, Greenwald J, Vega M, Castronovo S, Riek R, Choe S. *NMR structure of Mistic, a membrane-integrating protein for membrane protein expression.* Science (New York, NY). 2005;**307:**1317,Äì21.

76. Prosser RS, Evanics F, Kitevski JL, Al-Abdul-Wahid MS. *Current Applications of Bicelles in NMR Studies of Membrane-Associated Amphiphiles and Proteins†,‡.* Biochemistry. 2006;**45:**8453-65.

77. Ruan K-H, Wijaya C, Cervantes V, Wu J. *Characterization of the prostaglandin H2 mimic: Binding to the purified human thromboxane A2 receptor in solution.* Archives of Biochemistry and Biophysics. 2008;**477:**396-403.

78. Zhang H, Kurisu G, Smith JL, Cramer WA. *A defined protein–detergent–lipid complex for crystallization of integral membrane proteins: The cytochrome b6f complex of oxygenic photosynthesis.* Proceedings of the National Academy of Sciences. 2003;**100:**5160-3.

79. Jidenko M, Nielsen RC, Sørensen TL-M, Møller JV, le Maire M, Nissen P, et al. *Crystallization of a mammalian membrane protein overexpressed in Saccharomyces cerevisiae.* Proceedings of the National Academy of Sciences of the United States of America. 2005;**102:**11687-91.

80. Long SB, Campbell EB, MacKinnon R. *Crystal Structure of a Mammalian Voltage-Dependent Shaker Family K+ Channel.* Science. 2005;**309:**897-903.

81. Denisov IG, Grinkova YV, Lazarides AA, Sligar SG. *Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size.* Journal of the American Chemical Society. 2004;**126:**3477-87.

82. Nath A, Atkins W, Sligar S. *Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins.* Biochemistry. 2007;**46:**2059-69.

83. Lyukmanova EN, Shenkarev ZO, Paramonov AS, Sobol AG, Ovchinnikova TV, Chupin VV, et al. *Lipid-protein nanoscale bilayers: a versatile medium for NMR investigations of membrane proteins and membrane-active peptides.* J Am Chem Soc. 2008;**130:**2140-1.

84. Hagn F, Etzkorn M, Raschle T, Wagner G. *Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins.* J Am Chem Soc. 2013;**135:**1919-25.

85. Etzkorn M, Raschle T, Hagn F, Gelev V, Rice AJ, Walz T, et al. *Cell-free expressed bacteriorhodopsin in different soluble membrane mimetics: biophysical properties and*

NMR accessibility. Structure. 2013;**21:**394-401.

86. Bayburt TH, Sligar SG. *Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers.* Protein Science : A Publication of the Protein Society. 2003;**12:**2476-81.

87. Leitz AJ, Bayburt TH, Barnakov AN, Springer BA, Sligar SG. *Functional reconstitution of Beta2-adrenergic receptors utilizing self-assembling Nanodisc technology.* BioTechniques. 2006;**40:**601-2, 4, 6, passim.

88. Orwick MC, Judge PJ, Procek J, Lindholm L, Graziadei A, Engel A, et al. *Detergent-free formation and physicochemical characterization of nanosized lipid-polymer complexes: Lipodisq.* Angew Chem Int Ed Engl. 2012;**51:**4653-7.

89. Orwick-Rydmark M, Lovett JE, Graziadei A, Lindholm L, Hicks MR, Watts A. *Detergent-free incorporation of a seven-transmembrane receptor protein into nanosized bilayer Lipodisq particles for functional and biophysical studies.* Nano letters. 2012;**12:**4687-92.

90. Gohon Y, Dahmane T, Ruigrok RW, Schuck P, Charvolin D, Rappaport F, et al. *Bacteriorhodopsin/amphipol complexes: structural and functional properties.* Biophys J. 2008;**94:**3523-37.

91. Popot JL, Berry EA, Charvolin D, Creuzenet C, Ebel C, Engelman DM, et al. *Amphipols: polymeric surfactants for membrane biology research.* Cell Mol Life Sci. 2003;**60:**1559-74.

92. Pocanschi CL, Dahmane T, Gohon Y, Rappaport F, Apell HJ, Kleinschmidt JH, et al. *Amphipathic polymers: tools to fold integral membrane proteins to their active form.* Biochemistry. 2006;**45:**13954-61.

93. Dahmane T, Damian M, Mary S, Popot JL, Baneres JL. *Amphipol-assisted in vitro folding of G protein-coupled receptors.* Biochemistry. 2009;**48:**6516-21.

94. Zoonens M, Catoire LJ, Giusti F, Popot JL. *NMR study of a membrane protein in detergent-free aqueous solution.* Proc Natl Acad Sci U S A. 2005;**102:**8893-8.

95. Kielec JM, Valentine KG, Wand AJ. *A Method for Solution NMR Structural Studies of Large Integral Membrane Proteins: Reverse Micelle Encapsulation.* Biochimica et biophysica acta. 2010;**1798:**150-60.

96. Caffrey M. *A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes.* Acta crystallographica Section F, Structural biology communications. 2015;**71:**3-18.

97. Tate C, Schertler G. *Engineering G protein-coupled receptors to facilitate their structure determination.* Curr Opin Struct Biol. 2009;**19:**386-95.

98. Tate CG. *A crystal clear solution for determining G-protein-coupled receptor structures.* Trends in Biochemical Sciences. 2012;**37:**343-52.

99. Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, et al. *Structure and dynamics of the M3 muscarinic acetylcholine receptor.* Nature. 2012;**482:**552-6.

100. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D. *Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy.* Nat Struct Mol Biol. 2010;**17:**768-74.

101. Nietlispach D. *NMR Studies of a 7-Helix Transmembrane Receptor Sensory Rhodopsin II.* 2010**:**1-23.

102. Reckel S, Gottstein D, Stehle J, Löhr F, Verhoefen M, Takeda M, et al. *Solution NMR Structure of Proteorhodopsin.* Angew Chem Int Ed Engl. 2011;**50:**11942-6.

103. Bieri M, Kwan AH, Mobli M, King GF, Mackay JP, Gooley PR. *Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination.* Febs j. 2011;**278:**704-15.

104. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP. *Macromolecular NMR spectroscopy for the non-spectroscopist.* Febs j. 2011;**278:**687-703.

105. Keeler J. *Understanding NMR spectroscopy*. Second ed: Wiley; 2010.

106. Bermel W, Felli IC, Gonnelli L, Kozminski W, Piai A, Pierattelli R, et al. *High-dimensionality 13C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins.* J Biomol NMR. 2013;**57:**353-61.

107. Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Kozminski W. *High-dimensional NMR spectra for structural studies of biomolecules.* Chemphyschem : a European journal of chemical physics and physical chemistry. 2013;**14:**3015-25.

108. Stanek J, Podbevsek P, Kozminski W, Plavec J, Cevec M. *4D Non-uniformly sampled C,C-NOESY experiment for sequential assignment of 13C, 15N-labeled RNAs.* J Biomol NMR. 2013;**57:**1-9.

109. Pervushin K, Riek R, Wider G, Wüthrich K. *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* Proc Natl Acad Sci USA. 1997;**94:**12366-71.

110. Pervushin K. *The use of TROSY for detection and suppression of conformational exchange NMR line broadening in biological macromolecules.* J Biomol NMR. 2001;**20:**275-85.

111. Tugarinov V, Kay L. *Methyl Groups as Probes of Structure and Dynamics in NMR Studies of High-Molecular-Weight Proteins.* ChemBioChem. 2005;**6:**1567-77.

112. Tugarinov V, Kay LE. *Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods.* J Am Chem Soc. 2003;**125:**13868-78.

113. Tugarinov V, Hwang PM, Kay LE. *Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins.* Annual review of biochemistry. 2004;**73:**107-46.

114. Grzesiek S, Bax A. *Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR.* J Am Chem Soc. 1992;**114:**6291-3.

115. Grzesiek S, Bax A. *Amino Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins.* Journal of Biomolecular NMR. 1993;**3:**185-204.

116. Sattler M, Schleucher J, Griesinger C. *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients.* Prog Nucl Magn Reson Spectrosc. 1999;**34:**93-158.

117. Schubert M, Labudde D, Leitner D, Oschkinat H, Schmieder P. *A modified strategy for sequence specific assignment of protein NMR spectra based on amino acid type selective experiments.* J Biomol NMR. 2005;**31:**115-28.

118. Shen Y, Delaglio F, Cornilescu G, Bax A. *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR. 2009;**44:**213-23.

119. Bax A, Clore M, Gronenborn AM. *^{1}H-1H Correlation via Isotropic Mixing of 13C Magnetization, a New Three-Dimensional Approach for Assigning 1H and 13C Spectra of 13C Enriched Proteins.* J Magn Reson. 1990;**88:**425-31.

120. Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE. *Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of 13C labelled proteins.* J Magn Reson. 1990;**87:**620-7.

121. Lange OF, Rossi P, Sgourakis NG, Song Y, Lee H-W, Aramini JM, et al. *Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples.* Proceedings of the National Academy of Sciences. 2012;**109:**10873-8.

122. Schmitz C, Vernon R, Otting G, Baker D, Huber T. *Protein Structure Determination from Pseudocontact Shifts Using ROSETTA.* Journal of Molecular Biology. 2012;**416:**668-77.

123. Columbus L, Lipfert J, Klock H, Millett I, Doniach S, Lesley SA. *Expression, purification, and characterization of Thermotoga maritima membrane proteins for structure determination.* Protein Sci. 2006;**15:**961-75.

124. Lipfert J, Columbus L, Chu VB, Lesley SA, Doniach S. *Size and shape of detergent micelles determined by small-angle X-ray scattering.* J Phys Chem B. 2007;**111:**12427-38.

125. Flotenmeyer M, Weiss H, Tribet C, Popot JL, Leonard K. *The use of amphipathic polymers for cryo electron microscopy of NADH:ubiquinone oxidoreductase (complex I).* Journal of microscopy. 2007;**227:**229-35.

126. Lange OF, Baker D. *Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation.* Proteins. 2012;**80:**884-95.

127. Clore GM. *Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation.* Protein Science. 2011;**20:**229-46.

128. Battiste JL, Wagner G. *Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data†.* Biochemistry. 2000;**39:**5355-65.

129. Chen K, Tjandra N. *The use of residual dipolar coupling in studying proteins by NMR.* Topics in current chemistry. 2012;**326:**47-67.

130. Prestegard JH, Mayer KL, Valafar H, Benison GC. Determination of Protein Backbone Structures from Residual Dipolar Couplings. In: Thomas LJ, editor. Methods in Enzymology: Academic Press; 2005. p. 175-209.

131. Fitzkee N, Bax A. *Facile measurement of 1H–15N residual dipolar couplings in larger perdeuterated proteins.* Journal of Biomolecular NMR. 2010;**48:**65-70.

132. Allegrozzi M, Bertini I, Janik MBL, Lee Y-M, Liu G, Luchinat C. *Lanthanide-Induced Pseudocontact Shifts for Solution Structure Refinements of Macromolecules in Shells up to 40 Å from the Metal Ion.* Journal of the American Chemical Society. 2000;**122:**4154-61.

133. Yagi H, Pilla Kala B, Maleckis A, Graham B, Huber T, Otting G. *Three-Dimensional Protein Fold Determination from Backbone Amide Pseudocontact Shifts Generated by Lanthanide Tags at Multiple Sites.* Structure. 2013;**21:**883-90.

134. Campos M, Francetic O, Nilges M. *Modeling pilus structures from sparse data.* Journal of Structural Biology. 2011;**173:**436-44.

135. Ganguly S, Weiner Brian E, Meiler J. *Membrane Protein Structure Determination using*

Paramagnetic Tags. Structure. 2011;**19:**441-3.

136. Karaca E, Bonvin AMJJ. *Advances in integrative modeling of biomolecular complexes.* Methods. 2013;**59:**372-81.
137. Oxenoid K, Chou JJ. *The structure of phospholamban pentamer reveals a channel-like architecture in membranes.* Proc Natl Acad Sci U S A. 2005;**102:**10870-5.
138. Van Horn W, Kim H, Ellis C, Hadziselimovic A, Sulistijo E, Karra M, et al. *Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase.* Science. 2009;**324:**1726-9.
139. Berardi MJ, Shih WM, Harrison SC, Chou JJ. *Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching.* Nature. 2011;**476:**109-13.
140. Jaremko L, Jaremko M, Giller K, Becker S, Zweckstetter M. *Structure of the mitochondrial translocator protein in complex with a diagnostic ligand.* Science. 2014;**343:**1363-6.
141. Dawson PE, Kent SB. *Synthesis of native proteins by chemical ligation.* Annual review of biochemistry. 2000;**69:**923-60.
142. Kent SB. *Total chemical synthesis of proteins.* Chemical Society reviews. 2009;**38:**338-51.
143. Shah NH, Muir TW. *Inteins: Nature's Gift to Protein Chemists.* Chemical science (Royal Society of Chemistry : 2010). 2014;**5:**446-61.
144. Muralidharan V, Muir TW. *Protein ligation: an enabling technology for the biophysical analysis of proteins.* Nature methods. 2006;**3:**429-38.
145. Watson RP, Christen MT, Ewald C, Bumbak F, Reichen C, Mihajlovic M, et al. *Spontaneous self-assembly of engineered armadillo repeat protein fragments into a folded structure.* Structure. 2014;**22:**985-95.
146. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P. *Optimal isotope labelling for NMR protein structure determinations.* Nature. 2006;**440:**52-7.
147. Respondek M, Madl T, Gobl C, Golser R, Zangger K. *Mapping the orientation of helices in micelle-bound peptides by paramagnetic relaxation waves.* J Am Chem Soc. 2007;**129:**5228-34.
148. Skinner S, Moshev M, Hass MS, Ubbink M. *PARAssign—paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts.* Journal of Biomolecular NMR. 2013;**55:**379-89.
149. Bax A, Kontaxis G, Tjandra N. *Dipolar couplings in macromolecular structure determination.* Methods Enzymol. 2001;**339:**127-74.
150. Prestegard JH, al-Hashimi HM, Tolman JR. *NMR structures of biomolecules using field oriented media and residual dipolar couplings.* Q Rev Biophys. 2000;**33:**371-424.
151. Otting G, Ruckert M, Levitt MH, Moshref A. *NMR experiments for the sign determination of homonuclear scalar and residual dipolar couplings.* J Biomol NMR. 2000;**16:**343-6.
152. Lorieau J, Yao L, Bax A. *Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins.* J Am Chem Soc. 2008;**130:**7536-7.
153. Popot JL, Engelman DM. *Membrane protein folding and oligomerization: the two-stage model.* Biochemistry. 1990;**29:**4031-7.
154. Cohen L, Arshava B, Kauffman S, Mathew E, Fracchiolla K, Ding F-X, et al. *Guided*

reconstitution of membrane protein fragments. Biopolymers. 2014;**102:**16-29.

155. Lee BK, Jung KS, Son C, Kim H, Verberkmoes NC, Arshava B, et al. *Affinity purification and characterization of a G-protein coupled receptor, Saccharomyces cerevisiae Ste2p.* Prot Expr Pur. 2007**:** .

2 Structural Characterization of Triple Transmembrane Domain Containing Fragments of a Yeast G Protein-Coupled Receptor in an Organic: Aqueous Environment by Solution-State NMR Spectroscopy

Katrina E. Fracchiolla[1,2], Leah S. Cohen[1], Boris Arshava[1], Martin Poms[3], Oliver Zerbe[3], Jeffrey M. Becker[4], Fred Naider[1*]

[1]Department of Chemistry, The College of Staten Island, City University of New York (CUNY), Staten Island, NY 10314; [2]Department of Biochemistry, The Graduate Center, CUNY, New York, NY, 10016; [3]Institute of Organic Chemistry, University of Zurich, Switzerland; [4]Department of Microbiology, University of Tennessee, Knoxville, TN 37996

Short Running title: Characterization of Triple TM Domains of a GPCR

*) Department of Chemistry, The College of Staten Island, CUNY, 2800 Victory Boulevard, Staten Island, NY 10314, USA. Tel.: +1 718 982 3896; fax: +1 718 982 3910. E-mail address: fred.naider@csi.cuny.edu (F. Naider). Fred Naider is the Leonard and Esther Kurtz Term Professor at the College of Staten Island, City University of New York

**) Invited Article for the Anniversary Issue of Journal of Peptide Science. It is dedicated to Murray Goodman in memory of his strong support for this journal as member of the editorial board (1995-2004)

2.1 Abstract

This report summarizes recent biophysical and protein expression experiments on polypeptides containing the N-terminus, the first, second and third transmembrane domains and the contiguous loops of the α-factor receptor Ste2p, a G protein-coupled receptor. The 131-residue polypeptide Ste2p(G31-R161), TM1-TM3 was investigated by solution NMR in trifluorethanol/water: TM1-TM3 contains helical transmembrane domains at the predicted locations, supported by continuous sets of medium-range NOEs. In addition, a short helix N-terminal to TM1 was detected, as well as a short helical stretch in the first extracellular loop. Two 161-residue polypeptides, [Ste2p(M1-R161), NT-TM1-TM3], that contain the entire N-terminal sequence, one with a single mutation, were directly expressed and isolated from *E. coli* in yields as high as 30 mg/L. Based on its increased stability, the L11P mutant will be used in future experiments to determine long-range interactions. The study demonstrated that 3-TM domains of a yeast GPCR can be produced in isotopically labeled form suitable for solution NMR studies. The quality of spectra is superior to data recorded in micelles and allows more rapid data analysis. No tertiary contacts have been determined, and if present, they are likely transient. This observation supports earlier studies by us that secondary structure was retained in smaller fragments, both in organic solvents and in detergent micelles, but that stable tertiary contacts may only be present when the protein is imbedded in lipids.

Key words: GPCR fragments, biosynthesis, isotopic labeling , solution-state NMR

Abbreviation List: G protein-coupled receptor, GPCR; transmembrane, TM; nuclear magnetic resonance, NMR; Ste2p(G31-T110), TM1-TM3; trifluoroethanol, TFE; Ste2p (M1-R161), NT-TM1-TM3; Luria Broth, LB; ampicillin, Amp; isopropyl β-D-1-thiogalactopyranoside, IPTG; reverse phase high performance liquid chromatography, RP-HPLC; Circular Dichroism, CD; sodium dodecyl sulfate, SDS; lysomyristoylphosphatidylglycerol, LMPG; N-terminus, NT; extracellular loop 1, EL1; paramagnetic relaxation enhancements, PREs; Ste2p (M1-R161, L11P), NT^{L11P}-TM1-TM3.

2.2 Introduction

In the early 1960s the laboratory of Professor Murray Goodman initiated a seminal series of studies using synthetic homo-oligopeptides to understand aspects of peptide and protein structure. These studies led to fundamental insights into the chain length dependence of the formation of α-helices[1]. The initial studies were significantly expanded by the Toniolo group to provide information on β-structure formation both in solution and in the solid state[2, 3]. Most importantly the use of synthetic peptides as surrogates to probe questions related to protein structure was established. Since these early investigations thousands of investigations have been conducted on the conformational preferences of carefully designed peptides leading to breakthroughs in our understanding of β-hairpin or β-turn formation and the assumption of structure by β-and γ-peptides[4-7].

In this report we present studies on folding of a GPCR using a 3-transmembrane containing peptide surrogate of the α-factor receptor from the yeast *Saccharomyces cerevisiae*. This report describes our most recent work on Ste2p peptide surrogates, which began in the early 1990s, some of which was recently reviewed[8]. Here we report conformational preferences of a 130-residue peptide corresponding to the first 3 TM domains of the Ste2p receptor and two of its loop regions encompassing residues 31-161 in an organic:aqueous medium that was used as a membrane mimic previously[9]. This work reflects an outgrowth of training and an approach that one of us (FN) received in the Goodman laboratory and it is dedicated to the memory of Dr. Goodman and to the myriad of scientists who were trained under his mentorship.

Structural characterization of G protein-coupled receptors is notoriously difficult due to the inherent properties of these receptors, with high-resolution crystallographic structural information available to date for slightly more than 20 of the ~1000 identified GPCRs[10-39]. Modifications of most of these proteins through introduction of conformation-stabilizing mutations resulting in higher melting temperatures, crystallization in presence of antibodies, truncation of flexible segments, and insertion of a crystallization-nucleation protein into the long and flexible third intracellular loop were necessary in order to facilitate stability and crystal packing interactions. Although crystallographic studies of GPCRs have provided fundamental information, most of those modifications compromise the activity of the proteins, and therefore sparse information on the dynamics of the protein has been obtained so far. Accordingly, nuclear magnetic resonance (NMR) spectroscopy investigations are a valuable complement to crystallographic analyses.

NMR investigations of GPCRs are hampered by a number of technical issues associated with the production of the large quantities of isotopically labeled receptor required for NMR in the expression hosts that were used by crystallographers, by the difficulty of establishing conditions that mimic the biologically relevant environment of the receptor while still providing good-quality spectra, by the tendency of these membrane proteins to aggregate, and by the large size of the protein/lipid complexes. To date, NMR investigations have been reported for heptahelical integral membrane proteins, the vasopressin receptor, the CXCR1 receptor, the CB2 receptor, the β2-adrenergic receptor, and sensory rhodopsin and proteorhodopsin[40-46]. The studies on the CB2 and β2-adrenergic receptors have focused on dynamics utilizing a mixture of NMR and molecular stimulations to better understand conformational changes[45, 46]. High-resolution solution state structures have been reported for the bacterial GPCR analogues sensory rhodopsin and proteorhodopsin[44, 47]. These proteins are somewhat smaller in size when compared to most GPCRs, less flexible, and hence more well-behaved. Solid-state NMR was used for the determination of the NMR structure of the mammalian CXCR1 receptor in phospholipid bilayers[41]. While solid state approaches hold promise for the future, an NMR structure for a full-length, mammalian GPCR based strictly on measured constraints has yet to be determined.

In an effort to overcome some of the difficulties associated with the NMR characterization of full-length GPCRs, several groups have focused on the characterization of *fragments* of GPCRs. Fragments are often easier to express in high yields, and the smaller number of residues leads to less crowded spectra. Our group studies the yeast α-factor receptor, Ste2p, a 431-residue peptide ligand receptor, which we are using as a model system for GPCR methods development. We have published the only solution structure for a GPCR fragment containing two TMs [TM1-TM2; Ste2p(G31-T110)] in LPPG micelles and in 2,2,2-trifluoroethanol (TFE):water mixtures [9, 50]. In both cases, the fragment is helical and forms a hairpin. However, the helical hairpin is more stable in LPPG and only transiently formed in TFE:water. The formation of a tertiary structure, even a transient tertiary structure, supports the hypothesis that large domains of a GPCR can fold independently of the remainder of the protein.

All X-ray structures of GPCRs show that every TM domain is in contact with at least two other TM domains. Therefore, we hypothesized that increasing the size of our Ste2p fragment to 3TM domains would increase the probability of forming tertiary contacts and potentially result in a more stable structure through increased mutual stabilization. As a result, we expanded our structural characterization to a 3TM containing fragment of Ste2p(G31-R161), TM1-TM3. This fragment

contains 131 residues of Ste2p, including 19 residues from the N-terminal domain, the first TM through the third TM with connecting loops and five residues of the second intracellular loop. Here we report details of a structure and dynamics study on Ste2p TM1-TM3 in 50% TFE:water. Recently, we showed that the addition of the first 30-residues of the Ste2p N-terminus increased expression and the stability of Ste2p TM1-TM2 in NMR preparations[8]. We will also report on the expression and biophysical characteristics of Ste2p (M1-R161) NT-TM1-TM3, which contains 161-residues of Ste2p including the entire N-terminal domain and the same TMs and loops as above.

2.3 Materials and Methods

2.3.1 Assignment of Sidechain Resonances

NMR backbone assignment of the TM1-TM3 fragment of Ste2p in TFE:water at 45°C was previously reported[48]. Side chain resonances were assigned using the HCCH-TOCSY[49, 50], HCCC(CO)NH [51], and (HM)CM(CGCBCA)NH and (HM)CM(CBCA)NH[52] experiments using NMRView 5[53] and CARA[54]. Briefly, Cα and Cβ annotations from the backbone assignments were confirmed in the HCCC(CO)NH spectra. The latter were also useful to obtain frequencies of the connected protons. Sidechain assignments of aliphatic resonances were then completed with the help of HCCH-TOCSY spectra starting from anchoring resonances in the 2D [^{13}C,^{1}H]-HSQC experiments. In general, the [^{13}C,^{1}H]-HSQC spectrum was very crowded, and assignment of sidechain resonances using the CA and CB chemical shifts was difficult. Assignments of methyl groups in the ILV-labelled sample was performed using experiments published by the Kay group[52, 55] that start on methyl protons and connect to amide moieties. Knowledge of methyl assignments then facilitated sidechain assignments via HCCH-TOCSY correlations form the methyl moieties. The spectra were acquired using either a three-channel Varian NMR-S 600 MHz NMR spectrometer (Varian NMR Instrument, Palo Alto, CA) with a z-axis pulsed-field-gradient and a Varian 5mm [^{1}H,^{15}N,^{13}C,^{2}D] cryo-probe at the College of Staten Island, a three-channel Bruker AV-700 700 MHz NMR spectrometer (Bruker, Billerica, MA) equipped with a CRYO TXI inverse triple resonance cryoprobe at the University of Zurich, or a four-channel Bruker 800 MHz NMR spectrometer (Bruker, Billerica, MA) equipped with a CRYO TCI triple resonance cryoprobe at the New York Structural Biology Center.

2.3.2 Confirmation of Secondary Structure Localization in Ste2p TM1-TM3 (G31-R161) using ^{15}N T_2 relaxation and H,D Amide Exchange

^{15}N T_2 relaxation experiments were performed on a 0.5 mM [^{15}N]-TM1-TM3 sample solubilized in 50% TFE-d_2:(water+0.1% TFA) (Sigma, St. Louis, MO). A series of eight [^{15}N,^{1}H]-HSQC-based CPMG experiments were performed with varying relaxation times of 0, 10, 30, 50, 70, 110, 150 and 210 ms. The data were processed and the rate analysis function of NMRView5 was used to calculate the relaxation time for each residue.

H,D amide exchange analysis was performed on a 0.5 mM [^{15}N]-TM1-TM3 sample obtained by dissolving lyophilized protein in a fully deuterated solvent 50% TFE-d3:(D_2O + 0.1% TFA-d1) (Sigma, St. Louis, MO). A series of [^{15}N,^{1}H]-HSQC experiments were measured at 40 minute intervals for a total of 5 hours. Additional spectra were collected daily with the final spectrum collected eight days after the original sample preparation. The rate analysis function of NMRView5 was used to calculate the exchange time for each assigned residue, resulting in a logarithmic plot of the exchange time vs. residue number (see Figure 2.1A).

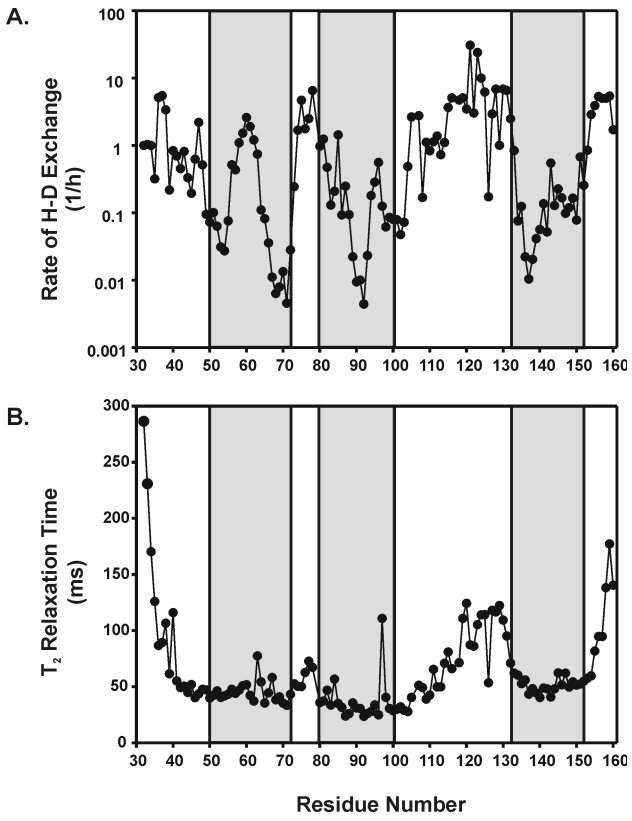


Figure 2.1: Evaluation of the secondary structure of Ste2p (G31-R161) by hydrogen-deuterium amide exhange and T_2 relaxation

(A) Exchange timeswere calculated from a series of [15N,1H]-HSQC experiments conducted at 45 °Cand plotted as a function of residue number with an inverse logarithmicscale. (B) The T2 relaxation times for all residues in TM1–TM3 weremeasured using the series of experiments conducted at 45 °C as described.The boxed helical boundaries are those calculated by sequence-based TMprediction software (average of four programs)

2.3.3 Assessment of the Relative Flexibility of Ste2p TM1-TM3 (G31-R161) using an Analysis of the 15N{1H}-NOE

A 0.5 mM sample of [^{15}N]-TM1-TM3 was solubilized in 50% TFE-d_2:water and subject to analysis of the ^{15}N{^{1}H}-NOE both at 45°C and 30°C. Amplitudes and volumes for each crosspeak in each data set were calculated using CARA. The ^{15}N{^{1}H}-NOE was derived by computing the ratio of peaks in the spectra with and without prior proton irradiation.

2.3.4 NOESY Assignment and Structure Calculation for Ste2p TM1-TM3(G31-R161)

Backbone chemical shift assignments were used to calculate torsion angles for TM1-TM3 using TALOS+[56]. The ATNOS-CANDID[57, 58] component of the UNIO suite, that is interfaced with the structure calculation program CYANA[59], was used for automated assignment of the 3D [^{13}C]-

and [^{15}N]-resolved NOESY spectra based on the assigned backbone and sidechain chemical shifts. A seven-cycle CYANA iteration using the 3D [^{15}N]- and [^{13}C]-resolved NOESY spectra as well as the TALOS-derived torsion angle restraints was used to calculate 80 structures, and the 20 lowest energy structures for TM1-TM3 were analyzed in detail.

2.3.5 Direct Expression of NT-TM1-TM3 Protein Fragments

A construct containing an N-terminal His_6-tag, followed by the full N-terminal tail and an S104C mutation (NT-TM1-TM3, Ste2p(M1-R161, S104C) corresponding to Ste2p(M1-R161) was cloned and expressed. The plasmid containing the NT-TM1-TM3 sequence was transformed into BL21(DE3), 20 µL of the transformation reaction was used to inoculate 50 mL of Luria Broth containing ampicillin (LB/Amp). The culture was grown overnight at 37°C and used to inoculate 1 L of LB/Amp, and the cells were grown at 37°C until an OD_{600} of approximately 1 was reached. The cells were then pelleted at 4000 rpm for 20 minutes and resuspended in M9 minimal medium using ^{13}C-glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources as needed, respectively (Cambridge Isotope Laboratories, Andover, MA). Expression was induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and the culture was incubated at 30°C for 22h, after which cells were pelleted at 5000 rpm for 20 min. Inclusion bodies were prepared from the cell pellets as previously described[60], with the addition of 5 mM dithiothreitol to the lysis buffer. Inclusion body pellets were dissolved in 70% TFA by sonication and purified by preparative reverse phase high performance liquid chromatography (RP-HPLC) on a Zorbax 300SB-C3 Prep HT (Agilent, Santa Clara, CA) 21.2 x 150 mm, 7 micron column at 60°C using gradient elution with an acetonitrile/2-propanol gradient from 30% to 72% Solvent B (80% acetonitrile, 20% 2-propanol, 0.1% TFA) where Solvent A contained 80% water, 20% 2-propanol, 0.1% TFA). After purification, ^{15}N, $^{15}N^{13}C$ and $^{15}N^{13}C^{2}H$-labelled proteins were lyophilized.

2.3.6 Circular Dichroism Spectroscopy of Ste2p TM1–TM3 and NT-TM1-TM3 Peptides

CD spectroscopy was used to access the secondary structure of the TM1–TM3 peptides. Experiments were carried out in both organic:aqueous solvents and detergents. Peptides were solubilized by sonication in 50% TFE:(water+0.1% TFA) and the concentration of the peptide stock solution was determined by UV absorbance at 280nm. The molar extinction coefficient of 12,950 $M^{-1}cm^{-1}$ was used for short M1M3 peptide and 19035 $M^{-1}cm^{-1}$ for NT-TM1-TM3 and NT^{L11P}-TM1-

TM3. The extinction coefficients were calculated using 1490 for Tyr, 5500 for Trp and 125 for Cys residues. This stock solution was portioned and lyophilized to yield approximately 20 μM or 7 μM solutions in detergent and TFE:water media, respectively. The detergents were prepared as 20 mM sodium dodecyl sulfate (SDS) (Sigma, St. Louis, MO) or lysomyristoylphosphatidylglycerol (LMPG) (Avanti Polar Lipids, Alabaster, Alabama) in 20 mM sodium phosphate buffer, pH 5.0. All samples were sonicated at 40°C for 5 min at 40 W using a S3000 sonicator (Farminigdale, NY) with a cup horn. The spectra were recorded on an Aviv Model 410 CD instrument (Aviv Biomedical, Lakewood, NJ). For the organic:aqueous studies, a quartz cuvette with a pathlength of 1 mm was used. For the detergent studies, a quartz cuvette with a pathlength of 0.2 mm was used. The spectra were collected over a wavelength range of 260 nm to 190 nm in increments of 1 nm. An average of four scans was taken in all cases, and the background subtraction used the spectrum from the solvent medium. The raw data were then converted in mean residue molar ellipticity (deg*cm^2*dmol^{-1}). Deconvolution analysis was performed using the CDNN software[61].

2.3.7 NMR Analysis to Determine Peptide Stability at High Concentrations

For NMR investigations in organic:aqueous media, 2-3 mg of lyophilized protein were solubilized in 175 μL of TFE and sonicated as above. After sonication, 175 μL of water containing 0.1% TFA were added and the sample was sonicated again. A clear solution was obtained and the sample was transferred to a Shigemi NMR tube (Shigemi, Allison Park, PA). ^{1}H NMR experiments were conducted at 45 °C on a three-channel Varian NMR-S 600 MHz NMR spectrometer (Varian NMR Instrument, Palo Alto, CA) with a z-axis pulsed-field-gradient and a Varian 5-mm [^{1}H,^{15}N,^{13}C,^{2}D] cryo-probe. Sample stability was assessed by measuring the overall peak integral for amide NH, NH_2 and aromatic area from 6.0 to 9.5 ppm for samples incubated for several days at 45°C.

2.4 Results and Discussion

2.4.1 Assignment of Side Chain Resonances of TM1-TM3

Side chain assignments were conducted using the HCCH-TOCSY, HCCC(CO)NH, (HM)CM(CGCBCA)NH, and (HM)CM(CBCA)NH experiments as described above. Data from all three experiments were combined to assign all [^{13}C,^{1}H]-HSQC crosspeaks. Complete side chain assignments were obtained except for the aromatic residues. A table of all chemical shift

assignments is provided in the Supplemental Materials (Supplemental Material Table S2.3) and chemical shifts have been added to the BMRB database under accession code 17211.

2.4.2 Assessment of the Secondary Structure of Ste2p TM1-TM3(G31-R161) by H,D Amide Exchange and ^{15}N T_2 Relaxation Experiments

H,D amide exchange and ^{15}N T_2 relaxation experiments were conducted in order to gain additional insight into the stability of secondary structure of the TM1-TM3 fragment in 50% TFE:water. Analysis of the H,D exchange results (Figure 2.1A) reveals that, in general, the residues in the predicted TM helices (boxed regions) displayed reduced H,D exchange rates when compared to those in the loop regions. The majority of the exchange times in the predicted helices range from approximately 10-200 h while they are between 0.02-1 h in the loop regions. Accelerated exchange is also observed in the middle of TM1, centered around the GXXXG motif. The predicted TM helix boundaries[48] seem to correlate very well with the exchange data for residues of TM1 and TM3, as slow exchange is only observed within the TM regions (indicated by boxes in Figure 2.1A), while fast exchange is restricted to loop regions. Except for residues in the very center of TM2, the exchange data for TM2 reveal that many residues near the ends of the helix possess exchange times between 1 and 10 h indicating reduced stability. This finding deviates from the conclusions previously reported that were derived from chemical shift analysis[48]. The putative TM3 displays more stable hydrogen bonds at its N-terminus, however the differences in exchange times across the helix are, in general, not as large as in TM2. In addition to the TM helices, the chemical shift analysis indicated the presence of helices encompassing residues 38 to 49 or 108 to 115 within the N-terminus (NT) and the extracellular loop 1 (EL1), respectively[48]. These helices are not detected in the H,D exchange data indicating that they are rather unstable.

^{15}N T_2 relaxation times (Figure 2.1B) are between 25 and 50 ms in the TM segments, and adopt values on the order of hundreds of milliseconds at the termini and around 50 ms in the loops. We observed increased relaxation times for the GXXXG region in TM1. TM1 again appears to be N-terminally extended, and this extension correlates well with the position of the N-terminal helix. Interestingly, relaxation times for residues in EL1 gradually increase until residue 120.

This gradual increase is consistent with the presence of a short helix in the beginning of EL1. Furthermore, the relaxation times in TM2 are the shortest of the 3 TM regions. Overall, the relaxation results seem to confirm the helix boundaries previously identified from secondary

chemical shifts[48]. We note that exchange data allows us to differentiate different degrees of helix stability for residues that may display very similar relaxation data as the latter in addition to helix stability may contain contributions from conformational exchange.

2.4.3 Assessment of the Relative Flexibility of Ste2p TM1-TM3(G31-R161) by Analysis of the 15N{1H}-NOE

Analysis of the $^{15}N\{^{1}H\}$-NOE has also been used to gain insights into the flexibility of membrane proteins[62]. Based on previous studies on the TM1-TM2 fragment of Ste2p[9], it was expected that the TM1-TM3 construct would be more structured at lower temperature. Accordingly, a series of HSQC experiments were run in 5°C decrements from 40°C to 25°C in order to transfer the backbone amide assignments (data not shown). It was found that 30°C was the lowest temperature at which chemical shift adaptations could be performed. Accordingly, $^{15}N\{^{1}H\}$-NOE experiments were performed at both 45°C and 30°C.

$^{15}N\{^{1}H\}$-NOE data collected at 45°C (Figure 2.2, top) closely reflect trends in the T_2 data. Both the N and C-termini display negative or small positive values, suggesting that these regions of the fragment are highly flexible. The three major regions with $^{15}N\{^{1}H\}$-NOE above 0.5 fall within the putative TM regions (dashed lines, Figure 2.2), but also include the short helix N-terminal to TM1. There is a small dip at the GXXXG motif in TM1. The putative loop regions show increased mobility relative to the TM regions. Again, the N-terminus of EL1 displays decreased mobility in agreement with the presence of some degree of order in that part of this loop.

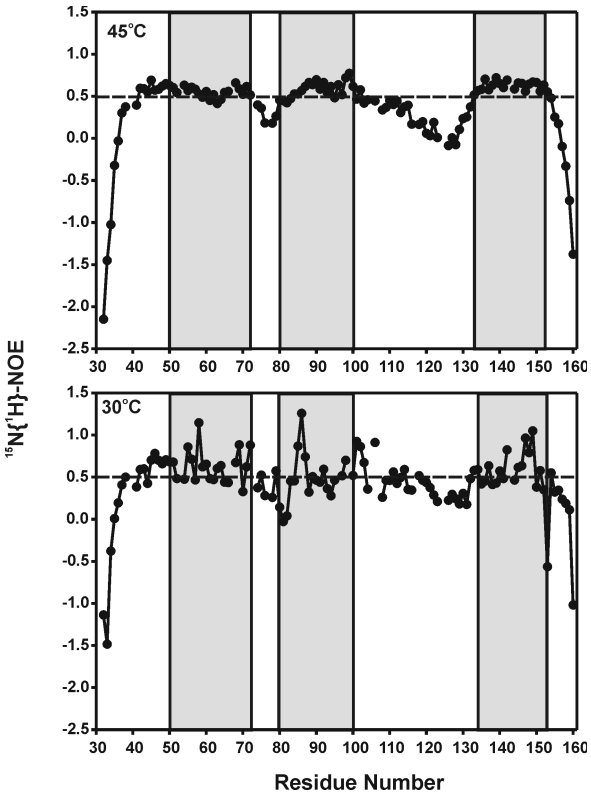


Figure 2.2: Evaluation of the relativemobility of the Ste2p TM1–TM3 (G31-R161) construct by ^{15}N{^{1}H}-NOE.

[15N,1H]-HSQC experiments were conducted at45 and 30 °C with and without amide proton irradiation. The ratio of thepeak amplitude with the pulse to the amplitude without the pulse forevery residue in the fragment was calculated and plotted as a functionof residue number. The boxed helical boundaries are those calculatedby sequence-based TM prediction software (average of four programs).

In general, data at 45°C seem to contain less noise in comparison to the 30°C data, likely due to the fact that peaks can be integrated more reliably at the higher temperature at which the lines are sharper and spectral overlap is less of a problem. At 30°C, the ^{15}N{^{1}H}-NOE for most regions of TM1-TM3 increases slightly. For example, TM1 appears to become less mobile at decreased temperature (many ratios >0.7), while the trend for increased mobility within the GXXXG motif is retained. We believe that, in general, the data supports increased overall rigidity for the protein at 30°C. The most significant differences are observed for EL1, for which the C-terminal half becomes rigidified as reflected in an increase of the ^{15}N{^{1}H}-NOE from 0 to approximately 0.25. We attribute the effect to a stabilization of secondary structure triggered by formation of tertiary contacts. The more frequent formation of these contacts results in part in exchange broadening of resonances, resulting in the observed problems in integration of spectra and deterioration of spectra quality at the lower temperature.

2.4.4 NOESY Assignment and Structure Calculation for Ste2p TM1-TM3(G31-R161)

A total of 342 intraresidual (|i-j|=0), 531 sequential (|i-j|=1), 873 short-range (|i-j| ≤ 1), 544 medium-range (1<|i-j|<5), and 5 long-range (|i-j|≥5) restraints were used in the calculation (Supplemental Material Figure S2.6 and Table S2.4). The twenty lowest-energy of the computed 80 structures were used to evaluate the structure. They had target function values ranging from 0.6 to 1.41. A total of 10 violations were observed (9 distance and 1 van der Waals violations). 90.3 % of dihedrals fell within the most favored regions of the Ramachandran plot, with the remaining 9.7% located in the additionally allowed regions (Supplemental Material Figure S2.7). As expected, the majority of the residues in the most favored region were in the right-handed α-helix portion of the plot.

An analysis of typical NOE contacts identified revealed a large number of *i, i+3* contacts throughout the N-terminus-TM1 region (Figure 2.3) for residues 36-49, 51-59, and 63-74. The absence of *i, i+4* contacts for residues 49-51, that corresponds to the junction between the N-terminal helix and TM1, suggests a destabilization of the α-helix at this position. Similarly, the absence of *i, i+4* contacts around the GXXXG motif in TM1 is consistent with reduced secondary chemical shifts, accelerated H,D exchange, and decreased T_2 relaxation.

The boundaries of the second helix are consistent with a C-terminal extension. This extension may also be the result of the helix-inducing properties of TFE:water, as it most likely represents an elongation of TM2 by part of EL1. Absence of *i, i+4* contacts for residues 102-105 indicates the transition from TM2 to EL1. Finally, the boundaries of TM3 are consistent with conclusions from sequence-based secondary structure prediction software[63-65].

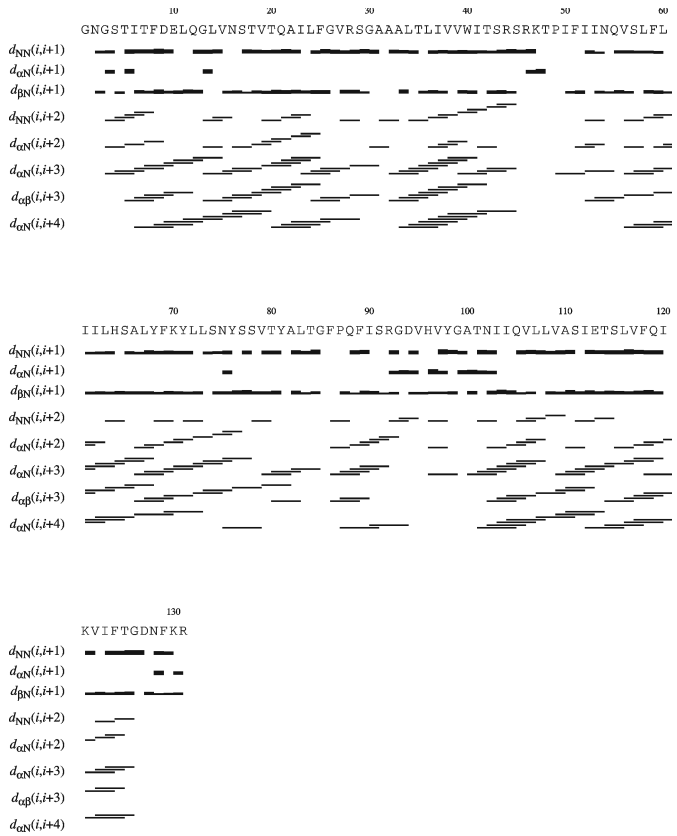


Figure 2.3: Interresidue NOE connectivities for the CYANA structure calculation of Ste2p TM1–TM3 (G31-R161).

The ATNOS-CANDID component of the UNIOsoftware suite was used for automatic assignment of the [13C]-resolved and [15N]-resolved NOESY experiments. Connectivities are displayed as a function of residue.

As very few long-range contacts were identified in the NOESY data, the final calculated conformers showed little convergence resulting in a large RMSD when superimposing the entire sequence. Therefore we superimposed only single TMs or even parts of them in the following analysis (Figure 2.4A, Table 2.1). The helices spanning the NT and TM1 (NT-TM1) and TM3 superimpose with backbone RMSDs of 3.02 for and 0.99 Å, respectively. As expected RMSDs improved when excluding segments for which no *i, i+4* contacts were observed (Figure 2.4B, Table 2.1). The resulting backbone RMSDs were 2.36 ± 1.16 Å for TM1, 1.03 ± 0.41 Å for TM2, 1.49 ± 0.51 Å for TM3, and 0.43 ± 0.18 Å and 0.69 ± 0.28 Å for the short NT and EL1 helices, respectively. The higher RMSD in TM1 is likely due to fluctuations about the central GXXXG motif. RMSD calculations performed using the TM boundaries from the sequence-based prediction software (Figure 2.4C, Table 2.1) and those derived from the homology model from Eilers[66] (Figure 2.4D, Table 2.1) resulted in the smallest RMSDs but these superimposed segments are also the shortest. Unfortunately, significant peak broadening at 25°C, did not allow recording of useful

spectra under these conditions. This precluded the detection of NOEs that would be expected if tertiary structure would be present at this temperature. The secondary structure determined based on NOEs is more accurately reflected in the T_2 data in comparison to H,D amide exchange.

Table 2.1: Calculated RMSD Values for secondary structure elements of TM1-TM3

Method	Putative Ste2p Region	Included Residues	Backbone RMSD	Heavy Atom RMSD
Structure[a]	NT-TM1	36-75	3.02 ± 1.41	3.81 ± 1.52
	TM2-EL1	83-111	2.09 ± 0.81	2.98 ± 0.84
	EL1	118-121	0.24 ± 0.11	1.17 ± 0.31
	TM3	132-152	0.99 ± 0.29	1.64 ± 0.29
NOE[b]	TM1	50-75	2.36 ± 1.16	3.05 ± 1.13
	TM2	86-103	1.03 ± 0.41	1.98 ± 0.46
	TM3	131-156	1.49 ± 0.51	2.07 ± 0.44
	NT	36-49	0.43 ± 0.18	1.31 ± 0.21
	EL1	105-115	0.69 ± 0.28	1.53 ± 0.40
Sequence-based	TM1	50-72	2.10 ± 1.09	2.77 ± 1.04
Calculation[c]	TM2	80-100	1.47 ± 0.5	2.5 ± 0.56
	TM3	132-152	0.99 ± 0.31	1.64 ± 0.29
	NT	38-49	0.39 ± 0.18	1.33 ± 0.23
	EL1	108-115	0.43 ± 0.22	1.14 ± 0.41
Eilers[d]	TM1	49-66	1.54 ± 0.79	2.1 ± 0.71
	TM2	84-100	0.82 ± 0.31	1.8 ± 0.34
	TM3	131-150	0.85 ± 0.28	1.49 ± 0.27
	NT	38-49	0.39 ± 0.18	1.33 ± 0.23
	EL1	108-115	0.43 ± 0.22	1.14 ± 0.41

[a] Boundaries determined in MOLMOL using the Kabsch-Sanders algorithm from the calculated structure.
[b]Boundaries from the calculated structure modified based on *i, i+4* NOE connectivites.
[c]Average boundaries from sequence-based TM prediction software.
[d]Boundaries from the rhodopsin-templated model of Ste2p

Our previous work has indicated that transient tertiary structure can be formed at lower temperatures (25°C) in an organic:aqueous solvent[9] but that more stable tertiary structures are formed in micellar environments[67]. Whether this also the case for TM1-TM3 is presently under investigation in our laboratories.

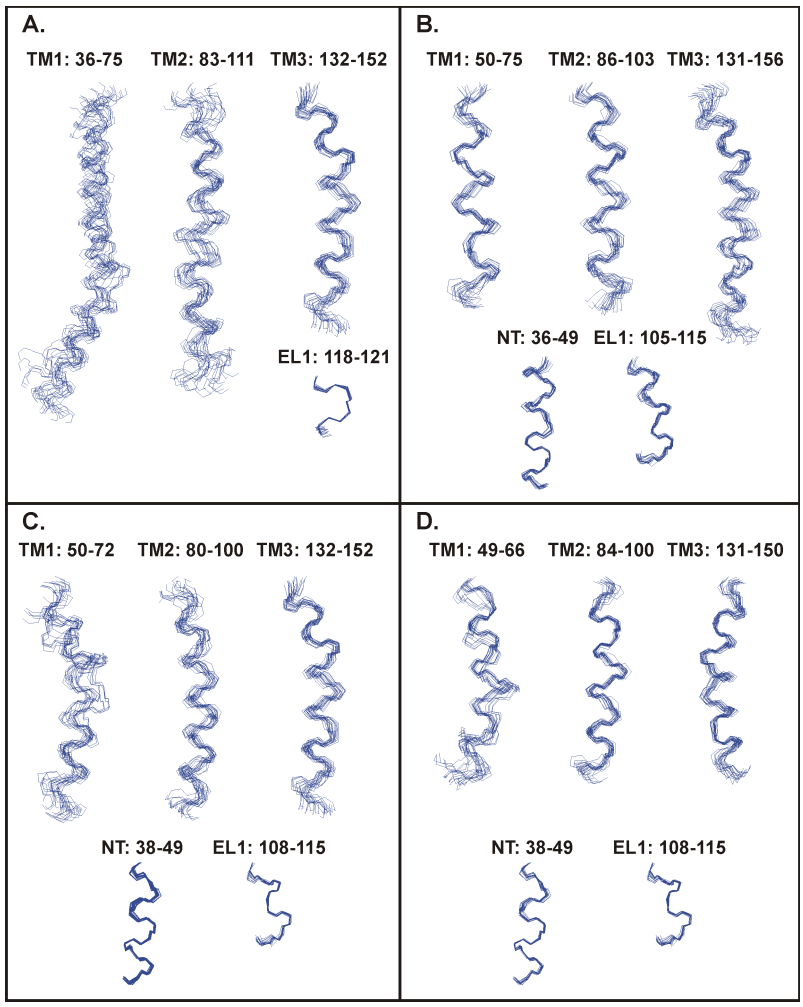


Figure 2.4: Convergence of the lowest 20 energy CYANA-calculated structures for Ste2p TM1–TM3 (G31-R161)

The helices were produced by fitting (A) thehelical boundaries calculated from the structure, (B) those determined from inter-residue NOE connectivities, (C) the calculated TM boundaries based onprediction software and the secondary shift boundaries for the NT and EL1 helices, and (D) the template-predicted TM boundaries and the secondary shiftboundaries for the NT and EL1 helices.

2.4.5 Direct Expression and Purification of NT-TM1-TM3: Maximizing Expression through Protein Engineering

Despite extensive attempts to optimize NMR conditions and obtain reproducibly spinlabeled 131-residue TM1-TM3 for measuring paramagnetic relaxation enhancements (PREs), the expression of the TM1-TM3(S104C) mutant proved extremely challenging. Direct expression led to very low and unreproducible yields, and CNBr cleavage of a TrpΔE fusion S104C protein resulted in no recoverable product. In contrast, following a recently developed approach for TM1-TM2[8], direct expression of the longer Ste2p(Met1-R161, S104C) [NT-TM1-TM3] without fusion to a

leader sequence (Supplementary Material Figure S2.8A) doubled yields to up to 30 mg per liter of culture. Expression was not always reproducible, however, and sometimes we experienced trouble with NMR sample preparation where the protein would go into solution but the solution would become gelatinous or the protein would precipitate over time. An additional mutation in the N-terminus, L11P (NT^{L11P}-TM1-TM3), resulted in more reproducible expression yields and stable protein samples (*vide infra*) (Supplementary Material Figure S2.8B). Previous work demonstrated that mutations in the N-terminus of Ste2p may result in changes in surface expression and signaling of this receptor *in vivo*[68-70]. SCAM analysis has indicated that downstream of residue 11 there is a β-sheet present[69, 70] and this region was also predicted to have β-strand potential between L8 and T16[71-73]. The presence of proline at position 11 can disrupt this conformation. Furthermore, recent work in the Becker lab on the functional implications of NT deletions revealed that Ste2p constructs missing residues 1-10 or 11-20 have increased biological function as indicated by reporter gene activity and/or mating efficiency (unpublished results). Since the N-terminus has been implicated in Ste2p dimerization[70] the destabilization of the β-strand by the Leu11Pro could have resulted in decreased aggregation resulting in more reliable expression yields and protein samples that will be amenable to NMR analysis. Our preliminary results demonstrate that this is the case for NT^{L11P}-TM1-TM3.

2.4.6 Biophysical characterization of NT-TM1-TM3 and comparison to TM1-TM3

CD spectroscopy was used to determine if the presence of the NT changes the overall secondary structure and fragment stability in TFE:water samples or micellar solutions. Previous analysis has indicated that a correlation between CD spectra and NMR analysis is useful to determine solvent conditions for NMR. The CD patterns of TM1-TM3, NT-TM1-TM3, and NT^{L11P}-TM1-TM3 were measured in two different TFE:water mixtures and micellar preparations. Furthermore, these samples were analyzed over a week to determine stability of the solution.

No difference in CD spectra was observed between 50% TFE:water and 67% TFE:water for the three protein samples TM1-TM3, NT-TM1-TM3, and NT^{L11P}-TM1-TM3 (Figure 2.5, left panels). The resulting spectra were analyzed for secondary structure tendencies using the deconvolution program CDNN[61] (Table 2.2). There was an overall decrease in helicity of the NT containing peptides which is likely due to an increase in the amount of β-sheet tendency in these peptides as described above. The β-strand percentage in NT^{L11P}-TM1-TM3 is decreased, presumably

due to the presence of Pro11 which should lower the β-strand tendency of the NT. As a result, the overall helicity of this peptide fragment increases when compared to the same peptide without the proline mutation. Similar structural tendencies are observed in micellar environments (Figure 2.5, right panels, Table 2.2). However, the amount of β-sheet structure in micellar environments compared to the organic aqueous mixtures is higher in all of the peptides. No differences were observed after one week at room temperature indicating that at the low concentration used in CD measurements, the protein samples are stable.

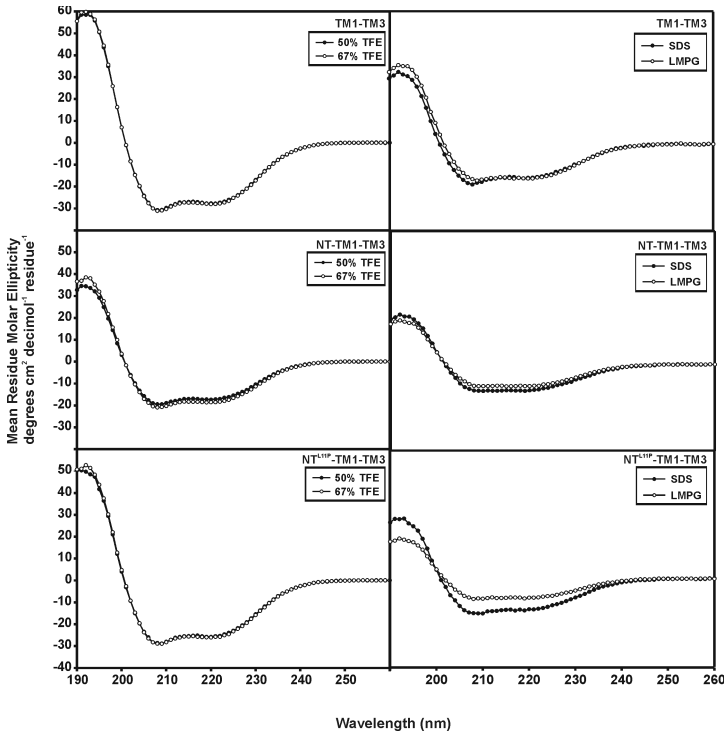


Figure 2.5: Secondary structure analysis of TM1–TM3, NT–TM1–TM3, and NT^{L11P}–TM1–TM3 using CD

Left panels: lyophilized peptideswere solubilized in 50%and 67% TFE as described. Right panels: lyophilized peptides were solubilized in 20mM SDS and LMPG as described. The presented spectra represent asolvent-subtracted average of four scans

Table 2.2: Percentages of secondary structures based on deconvolution of CD data using CDNN[64]

Peptide	TM1-TM3		NT-TM1-TM3		NT^{L11P}-TM1-TM3	
	α-helix	β-sheet	α-helix	β-sheet	α-helix	β-sheet
50% TFE	86.0%	1.6%	56.2%	6.1%	80.0%	2.2%
67% TFE	86.5%	1.5%	61.1%	5.1%	81.2%	2.1%
SDS	53.6%	6.8%	38.8%	12.7%	46.7%	9.1%

LMPG	55.3%	6.4%	34.2%	15.9%	32.0%	18.3%

Preliminary investigations of NMR sample conditions have shown that the NT-TM1-TM3 peptides are more stable in 67% TFE:water than in 50% TFE:water. Under similar conditions and concentrations the sample of NT^{L11P}-TM1-TM3 remains clear for longer time than TM1-TM3 or NT-TM1-TM3. After 6 days of incubation at 45°C, the best behaving sample of NT-TM1-TM3 lost ~30% of the peak intensity in both the NH/aromatics region and the aliphatic region. Conversely, the NT^{L11P}-TM1-TM3 sample retained 86% intensities in these same regions after one week. Most important is our observation that the NT^{L11P}-TM1-TM3 sample could be reproducibly prepared and remained NMR-stable for at least one week. This type of stability would allow for measurements of the 3D and 4D NMR experiments. This was not the case for NT-TM1-TM3 lacking Pro11. Pro11 may decrease the aggregation propensity of the N-terminus leading to a sample that is more suitable for NMR analysis. This observation may have implications for the development of a full-length Ste2p construct for NMR investigations.

2.5 Conclusions

Extensive NMR characterization of the TM1-TM3 fragment of Ste2p has been conducted in 50% TFE:water. Complete backbone and sidechain assignments have been made for data collected at 45°C. The backbone assignments were used for qualitative secondary structure and dynamics analysis via chemical shift analysis, as well as for H,D amide exchange, ^{15}N T_2 relaxation, and $^{15}N\{^1H\}$-NOE measurements. All analyses indicate the presence of three TM helices, with boundaries in close agreement with sequence-based predictions, with additional helicity observed in the N-terminal region and the N-terminus of EL1. Analysis of the $^{15}N\{^1H\}$-NOE suggests that certain areas of the fragment become less mobile at reduced temperatures. Reduction in the relative mobility of EL1 at 30°C would be consistent with the conclusion that the fragment adopts some degree of tertiary structure at reduced temperature. As 30°C is the optimal growth temperature for *Saccharomyces cerevisiae*, the organism expressing Ste2p, it is tempting to speculate that these tertiary structures are biologically relevant and formed during folding of the receptor.

The NOE-restrained structure calculation using data collected at 45°C reveals the presence of 4 helices. Three of them are in good agreement with the homology and secondary chemical shift predictions. Alignment of the calculated boundaries for the TM regions and the secondary shift

boundaries for the NT and EL1 helices reveals acceptable convergence and relatively low RMSD values. PRE experiments will need to be performed to identify possible long-range contacts. However, our preliminary results (data not shown) have not revealed such contacts in the organic:aqueous medium.

Our initial hypothesis was that increasing the number of TMs in a GPCR fragment would increase the probability for formation of interhelical contacts resulting in a better defined tertiary structure in TFE:water. Comparison of the TM1-TM2 and TM1-TM3 structural analyses in TFE:water suggests that additional work is necessary to test this hypothesis. TM1-TM2 adopted a transient tertiary structure with observable tertiary contacts as judged by PRE experiments. However, to date we have not been able to run reproducible PRE experiments with the analogous TM1-TM3 construct. As the size of the fragment has increased, the degree of difficulty of the NMR analyses has also increased. To a large extent this may be due to the tendency of the larger fragments to aggregate in membrane mimetic media, even in organic:aqueous mixtures. It may also be possible that TM2 and TM3 compete for overlapping helix association sites on TM1. This is likely not the case within the entire receptor, but in the fragments part of the interhelical contacts are missing, and hence alternative association modes may occur. As we have recently suggested in review articles, the NMR analysis of GPCRs is complicated by both conformational exchange broadening and intermolecular interactions[8, 74]. Part of this conformational flexibility is related to the mode of receptor activation, and the fact that most receptors display basal activity in absence of agonists. GPCRs often have regions with β-strand potential, which would create surfaces that can lead to aggregation at the high concentrations used in NMR anlaysis. Furthermore numerous studies have concluded that GPCRs form dimers and higher protomers in membranes (as reviewed in[75-77]). Our initial work with the Leu11Pro mutant of NT-TM1-TM3 suggests that properly engineered constructs may lower the tendency to aggregate which will lead to more stable NMR sample preparations and increased spectral quality for NMR analyses.

Acknowledgements

This work was supported by research grants GM22086 (F.N.) and GM22087 (J.M.B.) from the National Institutes of Health. Professor Naider is a member of the New York Structural Biology Center, which is a STAR center supported by the New York State Office of Science, Technology, and Academic Research.

2.6 References

1. Goodman M, Verdini AS, Toniolo C, Phillips WD, Bovey FA. *Sensitive criteria for the critical size for helix formation in oligopeptides.* Proc Natl Acad Sci U S A. 1969;**64:**444-50.
2. Goodman M, Naider F, Toniolo C. *Circular dichroism studies of isoleucine oligopeptides in solution.* Biopolymers. 1971;**10:**1719-30.
3. Toniolo C, Benedetti E. *Intramolecularly Hydrogen-Bonded Peptide Conformation.* Critical Reviews in Biochemistry and Molecular Biology. 1980;**9:**1-44.
4. Gellman SH. *Minimal model systems for beta sheet secondary structure in proteins.* Current opinion in chemical biology. 1998;**2:**717-25.
5. Cheng RP, Gellman SH, DeGrado WF. *beta-Peptides: from structure to function.* Chem Rev. 2001;**101:**3219-32.
6. Searle MS. *Insights into stabilizing weak interactions in designed peptide beta-hairpins.* Biopolymers. 2004;**76:**185-95.
7. Searle MS, Ciani B. *Design of beta-sheet systems for understanding the thermodynamics and kinetics of protein folding.* Curr Opin Struct Biol. 2004;**14:**458-64.
8. Cohen L, Arshava B, Kauffman S, Mathew E, Fracchiolla K, Ding F-X, et al. *Guided reconstitution of membrane protein fragments.* Biopolymers. 2014;**102:**16-29.
9. Cohen L, Arshava B, Neumoin A, Becker J, G5 untert P, Zerbe O, et al. *Comparative NMR analysis of an 80-residue G protein-coupled receptor fragment in two membrane mimetic environments.* Biochimica Et Biophysica Acta. 2011;**1808:**2674-84.
10. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science. 2000;**289:**739-45.
11. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, et al. *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science. 2007;**318:**1258-65.
12. Rasmussen S, Choi H, Rosenbaum D, Kobilka T, Thian F, Edwards P, et al. *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature. 2007;**450:**383-7.
13. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, et al. *The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.* Science. 2008;**322:**1211-7.
14. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP. *Crystal structure of the ligand-free G-protein-coupled receptor opsin.* Nature. 2008;**454:**183-7.
15. Scheerer P, Park J, Hildebrand P, Kim Y, Krauss N, Choe H, et al. *Crystal structure of opsin in its G-protein-interacting conformation.* Nature. 2008;**455:**497-502.
16. Warne T, Serrano-Vega M, Baker J, Moukhametzianov R, Edwards P, Henderson R, et al. *Structure of a beta1-adrenergic G-protein-coupled receptor.* Nature. 2008;**454:**486-91.
17. Chien EYT, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, et al. *Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist.* Science. 2010;**330:**1091-5.

18. Zhang C, Srinivasan Y, Arlow D, Fung J, Palmer D, Zheng Y, et al. *High-resolution crystal structure of human protease-activated receptor 1.* Nature. 2012;**492:**387-92.

19. Wu B, Chien E, Mol C, Fenalti G, Liu W, Katritch V, et al. *Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists.* Science. 2010;**330:**1066-71.

20. Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, et al. *Structure of the Adenosine A(2A) Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine.* Structure. 2011;**19:**1283-93.

21. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, et al. *Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation.* Nature. 2011;**474:**521-5.

22. Rasmussen S, Choi H, Fung J, Pardon E, Casarosa P, Chae P, et al. *Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor.* Nature. 2011;**469:**175-80.

23. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, et al. *Structure of the human histamine H1 receptor complex with doxepin.* Nature. 2011;**475:**65-70.

24. Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, et al. *Structure of the δ-opioid receptor bound to naltrindole.* Nature. 2012;**485:**400-4.

25. Haga K, Kruse A, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, et al. *Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist.* Nature. 2012;**482:**547-51.

26. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, et al. *Crystal structure of a lipid G protein-coupled receptor.* Science. 2012;**335:**851-5.

27. Kruse A, Li J, Hu J, Kobilka B, Wess J. *Novel insights into m3 muscarinic acetylcholine receptor physiology and structure.* J Mol Neurosci. 2014;**53:**316-23.

28. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, et al. *Crystal structure of the μ-opioid receptor bound to a morphinan antagonist.* Nature. 2012**:**1-7.

29. Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, et al. *Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic.* Nature. 2012;**485:**395-9.

30. White J, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, et al. *Structure of the agonist-bound neurotensin receptor.* Nature. 2012;**490:**508-13.

31. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, et al. *Structure of the human κ-opioid receptor in complex with JDTic.* Nature. 2012;**485:**327-32.

32. Huang J, Chen S, Zhang JJ, Huang X-Y. *Crystal structure of oligomeric β1-adrenergic G protein–coupled receptors in ligand-free basal state.* Nat Struct Mol Biol. 2013;**20:**419-25.

33. Hollenstein K, Kean J, Bortolato A, Cheng RKY, Doré AS, Jazayeri A, et al. *Structure of class B GPCR corticotropin-releasing factor receptor 1.* Nature. 2013;**499:**438-43.

34. Siu FY, He M, De Graaf C, Han GW, Yang D, Zhang Z, et al. *Structure of the human glucagon class B G-protein-coupled receptor.* Nature. 2013;**499:**444-9.

35. Wang C, Wu H, Katritch V, Han GW, Huang X-P, Liu W, et al. *Structure of the human smoothened receptor 7TM bound to an antitumor agent.* Nature. 2013;**497:**338-43.

36. Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, et al. *Structure of*

class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature. 2014;**511:**557-62.

37. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, et al. *Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator.* Science. 2014;**344:**58-64.

38. Zhang J, Zhang K, Gao Z, Paoletta S, Zhang D, Han G, et al. *Agonist-bound structure of the human P2Y12 receptor.* Nature. 2014;**509:**119-22.

39. Zhang K, Zhang J, Gao Z, Zhang D, Zhu L, Han G, et al. *Structure of the human P2Y12 receptor in complex with an antithrombotic drug.* Nature. 2014;**509:**115-8.

40. Tian C, Breyer RM, Kim HJ, Karra MD, Friedman DB, Karpay A, et al. *Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor.* J Am Chem Soc. 2005;**127:**8010-1.

41. Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, et al. *Structure of the chemokine receptor CXCR1 in phospholipid bilayers.* Nature. 2012;**491:**779-83.

42. Etzkorn M, Martell S, Andronesi OC, Seidel K, Engelhard M, Baldus M. *Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy.* Angew Chem Int Ed Engl. 2007;**46:**459-62.

43. Gautier A, Kirkpatrick J, Nietlispach D. *Solution-State NMR Spectroscopy of a Seven-Helix Transmembrane Protein Receptor: Backbone Assignment, Secondary Structure, and Dynamics.* Angew Chem Int Ed Engl. 2008**:** .

44. Reckel S, Gottstein D, Stehle J, Löhr F, Verhoefen M, Takeda M, et al. *Solution NMR Structure of Proteorhodopsin.* Angew Chem Int Ed Engl. 2011;**50:**11942-6.

45. Nygaard R, Zou Y, Dror R, Mildorf T, Arlow D, Manglik A, et al. *The dynamic process of beta(2)-adrenergic receptor activation.* Cell. 2013;**152:**532-42.

46. Kimura T, Vukoti K, Lynch DL, Hurst DP, Grossfield A, Pitman MC, et al. *Global fold of human cannabinoid type 2 receptor probed by solid-state 13C-, 15N-MAS NMR and molecular dynamics simulations.* Proteins. 2014;**82:**452-65.

47. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D. *Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy.* Nat Struct Mol Biol. 2010;**17:**768-74.

48. Caroccia K, Estephan R, Cohen L, Arshava B, Hauser M, Zerbe O, et al. *Expression and biophysical analysis of a triple-transmembrane domain-containing fragment from a yeast G protein-coupled receptor.* Biopolymers. 2011;**1808:**2674-84.

49. Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE. *Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of 13C labelled proteins.* J Magn Reson. 1990;**87:**620-7.

50. Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD. *A gradient-enhanced HCCH-TOCSY experiment for recording side-chan 1H and 13C correlations in H2O samples of proteins.* J Magn Reson Ser B. 1993;**101:**333-7.

51. Montelione GT, Emerson SD, Lyons BA. *A general approach for determining scalar coupling constants in polypeptides and proteins.* Biopolymers. 1992;**32:**327-34.

52. Tugarinov V, Kay LE. *Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods.* J Am Chem Soc.

2003;**125:**13868-78.

53. Johnson BA, Blevins RA. *NMR View: A computer program for the visualization and analysis of NMR data.* J Biomol NMR. 1994;**4:**603-14.

54. Keller R. *The Computer Aided Resonance Assignment.* 1252.

55. Tugarinov V, Hwang PM, Kay LE. *Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins.* Annual review of biochemistry. 2004;**73:**107-46.

56. Shen Y, Delaglio F, Cornilescu G, Bax A. *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR. 2009;**44:**213-23.

57. Herrmann T, Güntert P, Wüthrich K. *Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.* J Mol Biol. 2002;**319:**209-27.

58. Herrmann T, Güntert P, Wüthrich K. *Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS.* J Biomol NMR. 2002;**24:**171-89.

59. Güntert P. *Automated NMR structure calculation with CYANA.* Methods Mol Biol. 2004;**278:**353-78.

60. Cohen LS, Arshava B, Estephan R, Englander J, Kim H, Hauser M, et al. *Expression and biophysical analysis of two double-transmembrane domain-containing fragments from a yeast G protein-coupled receptor.* Biopolymers. 2008;**90:**117-30.

61. Bohm G, Muhr R, Jaenicke R. *Quantitative analysis of protein far UV circular dichroism spectra by neural networks.* Protein engineering. 1992;**5:**191-5.

62. Noggle J, Schirmer R. *The Nuclear Overhauser Effect - Chemical Applications.* 1252.

63. Hofman K, Stoffel W. *TMBASE: a database of membrane spanning protein segments*1993.

64. Hirokawa T, Boon-Chieng S, Mitaku S. *SOSUI: classification and secondary structure prediction system for membrane proteins.* Bioinformatics (Oxford, England). 1998;**14:**378-9.

65. Juretic D, Zoranic L, Zucic D. *Basic charge clusters and predictions of membrane protein topology.* Journal of chemical information and computer sciences. 2002;**42:**620-32.

66. Eilers M, Hornak V, Smith S, Konopka J. *Comparison of class A and D G protein-coupled receptors: common features in structure and activation.* Biochemistry. 2005;**44:**8959-75.

67. Neumoin A, Cohen L, Arshava B, Tantry S, Becker J, Zerbe O, et al. *Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles.* Biophys J. 2009;**96:**3187-96.

68. Shi C, Kaminskyj S, Caldwell S, Loewen M. *A role for a complex between activated G protein-coupled receptors in yeast cellular mating.* Proc Natl Acad Sci U S A. 2007;**104:**5395-400.

69. Shi C, Kendall SC, Grote E, Kaminskyj S, Loewen MC. *N-terminal residues of the yeast pheromone receptor, Ste2p, mediate mating events independently of G1-arrest signaling.* J Cell Biochem. 2009;**107:**630-8.

70. Uddin MS, Kim H, Deyo A, Naider F, Becker JM. *Identification of residues involved in homodimer formation located within a beta-strand region of the N-terminus of a Yeast G protein-coupled receptor.* Journal of receptor and signal transduction research.

2012;**32:**65-75.

71. Chou PY, Fasman GD. *Prediction of protein conformation.* Biochemistry. 1974;**13:**222-45.
72. Chou PY, Fasman GD. *Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins.* Biochemistry. 1974;**13:**211-22.
73. Cole C, Barber JD, Barton GJ. *The Jpred 3 secondary structure prediction server.* Nucleic acids research. 2008;**36:**W197-201.
74. Zerbe O. *First Solution Structures of Seven-Transmembrane Helical Proteins.* Angew Chem Int Ed Engl. 2011;**51:**860-1.
75. Gurevich VV, Gurevich EV. *GPCR monomers and oligomers: it takes all kinds.* Trends in neurosciences. 2008;**31:**74-81.
76. Fanelli F, Felline A. *Dimerization and ligand binding affect the structure network of A(2A) adenosine receptor.* Biochim Biophys Acta. 2011;**1808:**1256-66.
77. Kasai RS, Kusumi A. *Single-molecule imaging revealed dynamic GPCR dimerization.* Curr Opin Cell Biol. 2014;**27:**78-86.

2.7 Supplemental Material

Structural Characterization of Triple Transmembrane Domain Containing Fragments of a Yeast G Protein-Coupled Receptor in an Organic:Aqueous Environment by Solution-State NMR Spectroscopy

Katrina E. Fracchiolla[1,2], Leah S. Cohen[1], Boris Arshava[1], Martin Poms[3], Oliver Zerbe[3], Jeffrey M. Becker[4], Fred Naider[1]

[1]Department of Chemistry, The College of Staten Island, City University of New York (CUNY), Staten Island, NY 10314; [2]Department of Biochemistry, The Graduate Center, CUNY, New York, NY, 10016; [3]Institute of Organic Chemistry, University of Zurich, Switzerland; [4]Department of Microbiology, University of Tennessee, Knoxville, TN 37996

Table S2.3: Chemical Shift Assignments for Ste2p TM1-TM3 (G31-R161) in 50% TFE:water at 45°C.

Residue #	Amino Acid	δ (HN)	δ (N)	δ (CO)	δ (CA)	δ (HA)	δ (CB)	δ (HB)	Other
31	G			179.38					
32	N	8.56	118.98	175.13	52.74	4.81	38.27	2.86	
33	G	8.44	109.49	174.06	45.15	3.98			
34	S	8.06	115.84	174.64	58.46	4.49	63.22	3.93	
35	T	7.96	116.29	174.64	62.72	4.28	69.00	4.28	δ CG 20.74, δ HG 1.22,
36	I	7.77	122.17	176.31	62.34	4.15	37.84	1.88	δ CG 27.46, 16.19, δ HG 1.56, 1.23, 0.87, δ CD 11.64, δ HD 0.86
37	T	7.62	115.99	175.25	63.90	4.08	69.04	4.23	δ CG 20.39, δ HG 1.20
38	F	8.10	121.43	176.78	60.51	4.30	38.23	3.16	
39	D	8.28	117.80	177.43	55.28	4.40	36.56	3.02, 2.86	
40	E	8.07	120.58	177.76	57.25	4.09	27.46	2.09	δ CG 32.38, δ HG 2.42, 2.58
41	L	8.11	121.50	178.09	57.40	4.08	41.23	1.70	δ CG 26.19, δ HG 1.68, δ CD 22.61, δ HD 0.87
42	Q	8.15	117.31	178.29	58.49	3.83	27.78	1.97	δ CG 33.35, δ HG 2.13, 2.21
43	G	7.89	106.76	175.98	46.36	3.88			
44	L	8.07	125.55	179.88	57.66	4.17	41.29	1.86, 1.37	δ CG 26.05 δ HG 1.43, δ CD 22.48, 22.90, δ HD 0.85, 0.87
45	V	8.54	122.80	177.56	66.55	3.59	30.86	2.12	δ CG 19.79, 21.14, δ HG 0.92, 1.00
46	N	8.45	117.90	178.29	56.12	4.44	38.05	2.93, 2.78	
47	S	8.36	116.73	175.41	61.52	4.03	62.45	3.59	
48	T	8.00	119.62	175.89	66.37	3.98	68.44	4.37	δ CG 19.81, δ HG 1.25
49	V	8.39	122.51	177.23	66.25	3.69	31.17	2.12	δ CG 19.90, 21.46, δ HG 0.96, 1.03
50	T	7.91	114.44	175.69	66.93	3.84	68.24	4.28	δ CG 20.10 δ HG 1.21
51	Q	7.81	119.00	177.37	59.03	3.96	28.01	2.19	δ CG 33.77, δ HG 2.44
52	A	7.96	122.85	180.44	54.90	4.16	16.90	1.57	
53	I	8.32	120.70	178.16	64.52	3.78	37.32	2.03	δ CG 28.28, 15.69 δ HG 1.16, 1.76, 0.91 δ CD 11.66, δ HD 0.81
54	L	8.29	120.07	179.13	57.50	4.08	40.93	1.85	δ CG 26.15, δ HG 1.44, δ CD 21.56, 23.89, δ HD 0.83
55	F	8.47	119.21	178.60	60.26	4.31	38.22	3.24	
56	G	8.35	110.24	175.74	46.58	3.74			
57	V	8.43	122.71	178.10	65.45	3.80	31.32	2.21	δ CG 19.94, 20.74, δ HG 0.94, 1.06
58	R	8.30	119.75	177.78	58.17	4.07	29.51	1.89	δ CG 26.47, δ HG 1.70, δ CD 42.59, δ HD 3.13
59	S	8.07	113.77	175.47	59.97	4.18	62.90	3.61, 3.82	
60	G	7.81	110.47	174.45	45.62	4.02			
61	A	8.08	123.77	179.43	54.06	4.14	17.13	1.43	
62	A	7.98	121.82	178.48	54.37	4.08	16.99	1.44	
63	A	7.69	120.31	178.59	54.36	4.06	17.13	1.46	
64	L	7.80	117.06	178.01	57.10	4.10	41.28	1.79	δ CG 26.12, δ HG 1.71, δ CD 22.57, 24.32, δ HD

									0.88, 0.92
65	T	7.73	113.30	174.99	66.43	3.80	68.14	4.29	δ CG 19.96, δ HG 1.25
66	L	7.60	120.89	178.45	57.78	3.98	40.92	1.80	δ CG 26.09, δ HG 1.55, δ CD 21.98, 23.03, δ HD 0.80, 0.85
67	I	7.82	119.70	177.52	64.90	3.84	37.34	2.10	δ CG 28.32, 16.26, δ HG 1.08, 1.00 δ CD 11.34 δ HD 0.82
68	V	8.19	120.37	180.00	66.80	3.69	30.91	2.28	δ CG 20.02, 21.57, δ HG 0.94, 1.07
69	V	8.58	124.21	178.26	66.84	3.55	30.80	2.30	δ CG 19.81, 21.67, δ HG 0.98, 1.07
70	W	8.74	124.55	178.65	60.99	4.27	28.26	3.44, 3.59	
71	I	9.13	119.26	179.26	64.95	3.57	37.57	1.88	δ CG 28.14, 16.01, δ HG 1.08, 0.94 δ CD 12.07 δ HD 0.89
72	T	8.48	114.65	176.71	65.66	3.94	68.53	4.28	δ CG 20.25 δ HG 1.25
73	S	8.28	117.76	175.58	60.87	4.56	62.68	3.93	
74	R	7.69	121.10	177.23	56.64	4.06	29.30	1.73	δ CG 26.07, δ HG 1.36, δ CD 42.15, δ HD 2.64
75	S	7.82	114.57	174.44	59.28	4.30	63.08	3.89	
76	R	7.68	120.81	175.65	56.16	4.22	29.04	1.80	δ CG 26.53, δ HG 1.59, δ CD 42.69 δ HD 3.13
77	K	7.82	119.66	175.49	55.40	4.36	32.22	1.73	δ CG 24.09, δ HG 1.39, δ CD 28.22, δ CE 41.65, δ HE 2.94
78	T	7.62	114.88	172.88	60.57	4.57	68.90	4.21	
79	P			176.80	63.75		30.68	1.81	δ CG 26.74, δ HG 2.00, δ CD 50.02, δ HD 3.82, 3.67
80	I	7.43	117.79	176.22	62.52	3.86	37.20	1.81	δ CG 27.49, 16.19, δ HG 1.22, 0.85, δ CD 11.44, δ HD 0.87
81	F	7.50	120.00	176.87	59.58	4.38	38.35	3.19	
82	I	7.44	119.79	176.82	63.18	3.83	37.25	1.90	δ CG 27.86, 16.35, δ HG 1.26 δ CD 11.24, δ HD 0.86, 0.86
83	I	7.83	120.15	178.18	63.68	3.75	37.15	1.91	δ CG 27.86, 15.84, δ HG 1.64 δ CD 11.56, δ HD 0.81
84	N	8.15	120.00	176.09	55.61	4.43	37.98	2.78	
85	Q	7.90	119.36	177.60	58.57	4.02	27.95	2.12	δ CG 33.89, δ HG 2.53
86	V	8.37	120.05	177.25	65.94	3.75	30.96	2.15	δ CG 19.86, 21.26, δ HG 0.91, 1.02
87	S	8.06	114.81	175.19	61.99	4.18	62.58		
88	L	7.68	121.80	177.90	57.35	4.07	40.87	1.74	δ CG 26.11, δ HG 1.43, δ CD 22.46, 23.01 δ HD 0.74, 0.85
89	F	8.00	119.33	176.80	60.47	4.25	38.04	3.30	
90	L	8.34	117.92	180.19	57.71	3.98	40.91	1.38	δ CG 26.22, δ HG 2.06, δ CD 21.44, 24.01, δ HD 0.91
91	I	8.18	123.14	179.12	64.94	3.75	37.38	2.15	δ CG 28.05, 15.77, δ HG 1.15, 0.90 δ CD 11.58, δ HD 0.83
92	I	8.48	125.11	178.22	65.18	3.63	36.90	2.00	δ CG 27.91, 15.20, δ HG 1.08, 0.87,

									δ CD 11.48, δ HD 0.84
93	L	8.66	121.37	178.61	57.62	3.96	40.97	1.37	δ CG 25.97, δ HG 1.46, δ CD 22.31, 22.92, δ HD 0.76, 0.69
94	H	8.25	114.46	176.39	59.30	4.16	27.18	3.33	
95	S	8.24	116.44	174.73	61.54	4.17	62.43	3.93	
96	A	8.41	125.77	179.71	54.84	4.21	17.01	1.53	
97	L	8.09	118.30	178.68	57.15	4.08	40.97	1.49	δ CG 26.06, δ HG 1.54, δ CD 22.60, 23.70, δ HD 0.86, 0.83
98	Y	8.01	120.29	177.07	60.54	4.30	37.15	3.25	
99	F	8.27	119.06	177.04	60.78	4.26	37.81	3.21	
100	K	8.14	118.74	178.95	59.26	3.90	31.08	1.98	δ CG 24.69, δ HG 1.61,1.38, δ CD 28.65, δ CE 41.59, δ HE 2.90
101	Y	8.22	122.40	177.81	60.43	4.14	37.16	3.10, 3.21	
102	L	8.38	122.87	179.12	57.86	3.79	41.05	1.71	δ CG 25.85, δ HG 1.46, δ CD 22.26, 23.77, δ HD 0.80, 0.74
103	L	8.37	119.14	179.72	56.96	3.98	40.88	1.54	δ CG 26.23, δ HG 1.58, δ CD 22.22, 23.93, δ HD 0.76, 0.83
104	S	8.06	115.42	175.44	60.69	4.19	62.54	3.96	
105	N	7.77	120.92	175.71	55.17	4.50	38.55	2.47, 2.51	
106	Y	8.32	120.70	177.40	59.82	4.29	37.44	3.08	
107	S	8.23	115.72	175.08	60.45	4.27	62.58	3.96	
108	S	7.79	117.67	175.48	60.53	4.30	62.47	4.01	
109	V	7.85	121.92	176.75	65.16	3.83	31.22	2.10	δ CG 19.93, 20.55, δ HG 0.94, 0.98
110	T	7.71	113.43	176.44	64.89		68.19	4.06	δ CG 20.41, δ HG 1.28
111	Y	7.88	122.51	177.11	59.99	4.29	37.27	3.09	
112	A	7.87	122.71	178.92	54.12	4.04	17.31	1.48	
113	L	7.94	116.72	178.08	56.11	4.23	41.83	1.80, 1.55	δ CG 26.16, δ HG 1.81, δ CD 21.53, 23.88, δ HD 0.80, 0.79
114	T	7.65	109.20	175.13	62.58	4.24	69.61	4.23	δ CG 20.19, δ HG 1.21
115	G	7.88	109.65	173.65	44.89	3.93			
116	F	7.88	121.28	173.81	58.56	4.67	37.72	3.10	
117	P			177.54	64.30		30.47	1.72	δ CG 27.06, δ HG 1.90, δ CD 49.63, δ HD 3.72, 3.50
118	Q	8.05	117.91	176.48	58.01	4.06	27.46	2.33	δ CG 33.17, δ HG 2.29
119	F	7.77	119.21	176.73	58.75	4.43	38.34	3.14	
120	I	7.70	119.06	176.57	61.92	3.93	37.45	1.78	δ CG 16.53, 26.96, δ HG 1.37, 1.12, 0.82, δ CD 11.49, δ HD 0.78
121	S	7.74	116.16	174.76	58.86	4.33	63.08	3.95	
122	R	7.75	121.21	176.58	56.41	4.25	29.56	1.81	δ CG 26.81, δ HG 1.81, δ CD 42.69, δ HD 3.14
123	G	7.99	107.47	173.74	44.99	3.87			
124	D	7.94	118.42	174.72	52.67	4.68	37.19	2.83, 2.91	
125	V	7.64	119.56	175.52	62.41	3.96	31.67	2.05	δ CG 19.66, 19.75, δ HG 0.86, 0.83
126	H	8.30	120.58	173.80	54.99	4.66	28.08	3.18, 3.12	
127	V	7.70	119.81	175.38	62.31	4.01	32.10	1.98	δ CG 19.57, 20.02, δ HG

									0.86, 0.84
128	Y	7.82	122.44	176.02	57.82	4.52	38.00	3.05, 2.93	
129	G	8.00	109.45	173.66	45.01	3.93			
130	A	7.83	123.44	177.67	52.83	4.30	18.02	1.41	
131	T	7.89	110.99	174.54	63.34	4.21	68.73	4.21	δ CG 20.54, δ HG 1.26
132	N	7.96	119.89	175.63	53.88	4.67	38.21	2.83	
133	I	7.72	120.34	176.06	62.54	3.94	37.38	1.91	δ CG 27.77, 16.21, δ HG 1.53, 0.91, δ CD 11.58, δ HD 0.86
134	I	7.66	120.35	176.84	63.66	3.75	36.66	1.92	δ CG 28.11, 16.01 δ HG 1.59, 0.92, δ CD 10.83, δ HD 0.83
135	Q	7.59	117.70	177.22	59.11	3.96	28.02	2.14	δ CG 33.84, δ HG 2.33
136	V	7.41	118.25	177.30	65.92	3.60	30.96	2.25	δ CG 19.79, 21.08, δ HG 0.93, 1.02
137	L	7.97	119.66	180.08	57.20	4.07	40.97	1.91	δ CG 25.81, δ HG 1.53, δ CD 21.32, 23.84, δ HD 0.80, 0.84
138	L	8.48	122.61	178.86	58.09	4.09	41.18	1.84	δ CG 26.24, δ HG 1.70, δ CD 22.37, 22.92, δ HD 0.84, 0.71
139	V	8.09	120.17	177.96	66.76	3.59	30.80	2.19	δ CG 19.85, 21.67, δ HG 0.93, 1.08
140	A	8.63	121.95	181.00	54.89	4.11	16.82	1.49	
141	S	8.33	116.94	176.08	62.12	4.20		3.95	
142	I	8.32	126.22	178.16	64.67	3.74	37.69	2.01	δ CG 28.24, 15.69, δ HG 1.30, 0.89 δ CD 11.72, δ HD 0.82
143	E	8.76	119.16	177.81	59.26	3.98	27.35	2.22, 2.11	δ CG 32.40, δ HG 2.37, 2.63
144	T	8.16	112.48	175.68	66.04	3.93	68.51	4.29	δ CG 20.17, δ HG 1.24
145	S	7.77	117.66	175.83	61.52	4.13	62.60	4.02	
146	L	8.13	124.30	178.18	57.74	4.20	41.19	1.89	δ CG 26.02, δ HG 1.66, δ CD 22.19, 23.26, δ HD 0.85, 0.87
147	V	7.99	118.37	177.52	66.19	3.63	30.86	2.18	δ CG 19.97, 21.46, δ HG 0.92, 1.06
148	F	8.06	119.53	176.76	60.92	4.19	38.33	3.24	
149	Q	7.97	117.49	177.63	58.87	3.89	28.35	2.33	δ CG 34.03, δ HG 2.64
150	I	8.22	119.44	177.35	64.40	3.74	37.22	1.99	δ CG 28.00, 15.40, δ HG 1.09, 0.66, δ CD 11.40, δ HD 0.81
151	K	8.15	119.48	179.24	59.21	3.99	31.46	1.97, 1.85	δ CG 24.51, δ HG 1.38, 1.57, δ CD 28.69, δ HD 1.63,δ CE 41.56, δ HE 2.90
152	V	8.05	120.96	177.95	65.88	3.59	30.82	2.06	δ CG 20.05, 21.08, δ HG 0.81, 1.02
153	I	7.98	121.40	178.12	64.14	3.70	37.14	1.87	δ CG 28.05, 15.84, δ HG 1.84, 0.66, δ CD 11.51, δ HD 0.82
154	F	8.50	118.73	177.46	59.69	4.42	38.23	3.17	
155	T	7.95	112.79	175.65	62.90	4.16	69.05	4.32	δ CG 20.84, δ HG 1.25
156	G	8.02	110.07	174.42	45.64	3.94			
157	D	8.12	118.47	175.00	53.16	4.62	37.03	2.86	
158	N	7.98	117.96	174.52	53.62	4.56	38.14	2.61, 2.58	
159	F	7.75	119.15	174.92	57.48	4.56	38.44	3.04,	

								3.16	
160	K	7.68	121.35	175.33	55.64	4.29	32.24	1.85	δ CG 23.87, δ HG, 1.39, δ CD 28.42, δ HD 1.73, δ CE 41.70, δ HE 2.98
161	R	7.83	121.68	177.13	54.40	4.37	30.04	1.93, 1.77	δ CG 26.41, δ HG 1.64, δ CD 42.69, δ HD 3.17

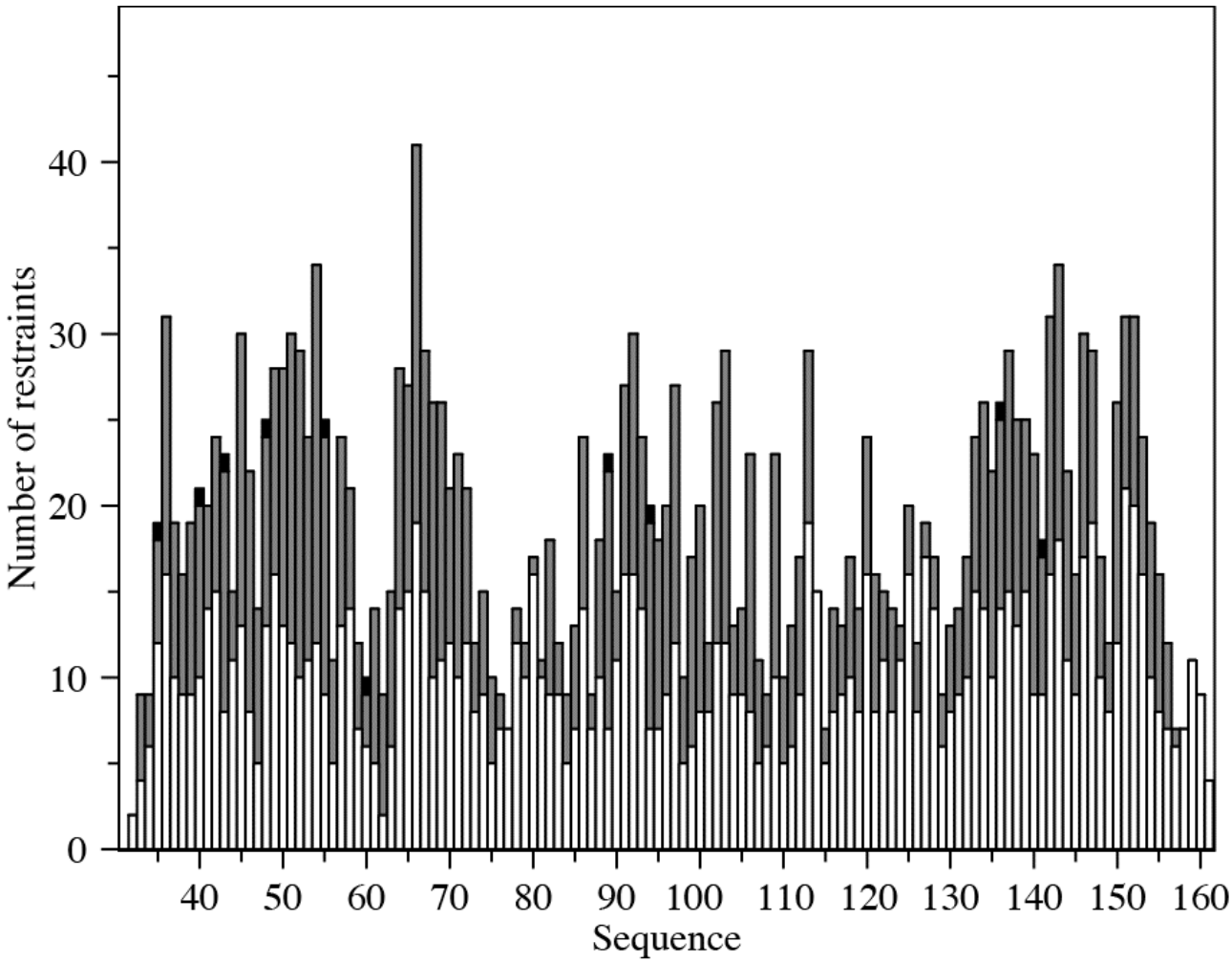


Figure S2.6: Number of restraints used for the CYANA structure calculation for Ste2p TM1-TM3 (G31-R161)

White bars indicate intra-residue connectivities. Light gray bars indicate short-range connectivities. Dark gray bars indicate medium-range connectivities. Black bars indicate long-range connectivities.

Table S2.4: NMR Constraints and Structural Statistics for 20 Structures of TM1-TM3 in 50% TFE:water

Distance Constraints	1422
Short Range $\|i-j\|\leq 1$	873
Medium Range $1<\|i-j\|<5$	544
Long Range $\|i-j\|\geq 5$	5
Hydrogen Bond Constraints	80
Dihedral Angle Constraints	240

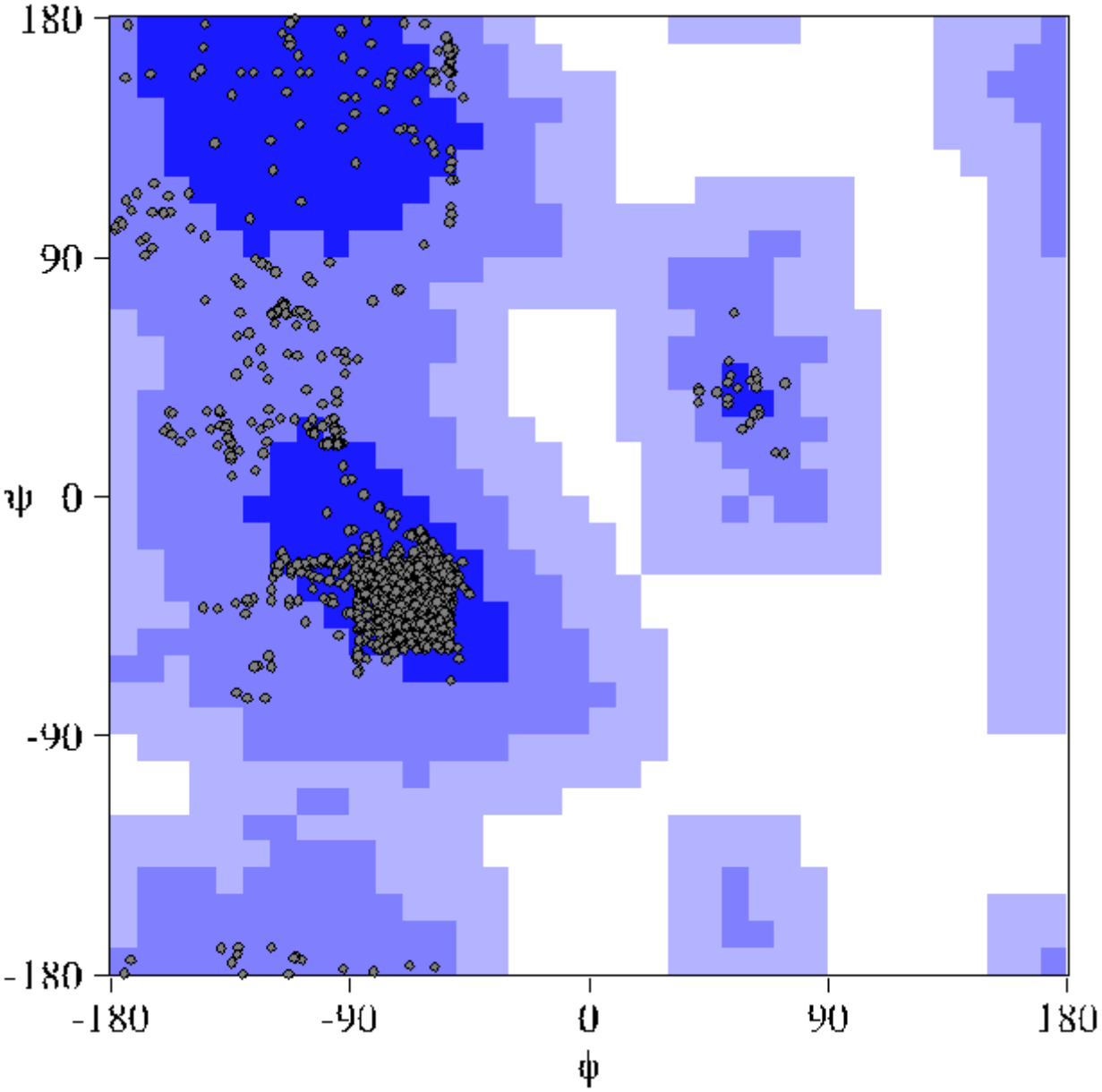


Figure S2.7: Ramachandran plot from CYANA structure calculation for Ste2p TM1-TM3 (G31-R161)

The majority of residues, 90.3%, are found within the most favored regions of the plot, with the remaining 9.7% in the additionally allowed regions. Most of the residues fall within the right-handed α-helix region of the plot.

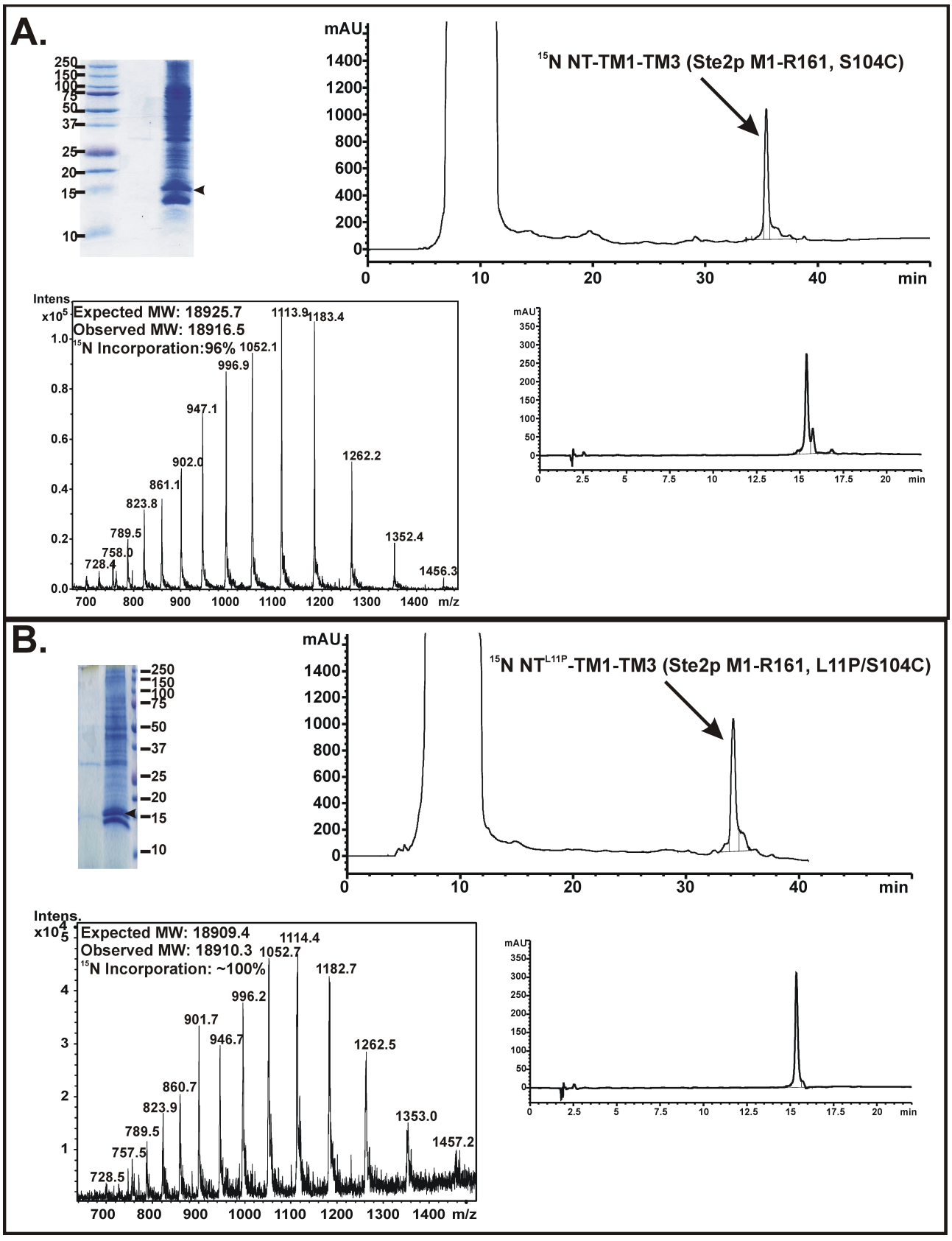


Figure S2.8: Expression and Purification of NT-TM1-TM3 peptides.

Large-scale expression of Ste2p NT-TM1-TM3 (M1-R161, S104C)(A) and Ste2p NTL11P-TM1-TM3 (M1-R161,L11P/S104C)(B). Expression was performed as described in the Materials and Methods. SDS-PAGE (gel, top left) was used to evaluate expression levels. Preparative HPLC purification of IB pellets revealed that the desired product eluted from the column at approximately 34-35 minutes (top right chromatogram in each box). ESI-MS analysis of the eluted product revealed that the polypeptides had the expected molecular weight with >96% [15N]-incorporation (bottom right spectra in each box). Analytical HPLC (bottom right chromatogram) indicated that the peptides were pure. The shoulder in the top analytical chromatogram is likely some conformational variant as judged by mass spectroscopy

3 Structural Models of Intermediates in GPCR Folding as Determined by NMR

M. Poms[1], (S. Jurt[1], Daniel Gottstein, K. Caroccia, L. Cohen, P. Güntert, F. Naider[2] and O. Zerbe[1]*)

[1]Department of Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057

[2]Department of Chemistry, The College of Staten Island, City University of New York (CUNY), Staten Island, NY 10314

*Corresponding author:
Oliver Zerbe
Department of Chemistry
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland
phone: +41-44-6354263
fax: +41-44-6356882
email: oliver.zerbe@chem.uzh.ch

3.1 Abstract

GPCRs are one of the most interesting classes of biomolecules to study as potential drug targets. As membrane proteins, containing seven transmembrane helices (TMs), structural characterization of GPCRs remains quite challenging. In this work, we describe the study of fragments from the yeast GPCR Ste2p using solution-state NMR spectroscopy. Therein, several constructs containing one or more TMs, namely TM1 (containing the first transmembrane helix), TM12, TM123 and TM127, a construct in which TM7 is covalently linked to TM2, are investigated to identify interhelical contacts. Systematic elongation of receptor fragments by the addition of further helices allows the investigation of early receptor folding. It can be observed that TM1, which is destabilized around a G-XXX-G motif, becomes rigidified upon the addition of TM2, as demonstrated in the previously solved TM12 fragment[1]. The addition of a further helix in case of the TM123 fragment, however, appears to destabilize the tertiary fold and the membrane integration due to unsatisfied polar residues in the center of TM3, which also makes TM123 prone to aggregation. The addition of TM7 yielded the much more stable TM127 fragment, which nonetheless seems to be highly dynamic. Making extensive use of sparse sampling NMR data, such as paramagnetic relaxation enhancement (PRE) and residual dipolar couplings (RDC), it could be demonstrated that interhelical contacts are formed, and that the helices are well embedded in the micelle. However, TM7 is performing a hinge motion about the central Leu-Pro-Leu motif, and the 3-TM bundle appears to be less compact when compared to a homology model.

KEYWORDS: GPCR, membrane proteins, structural biology, Ste2p receptor

3.2 Introduction

G-protein coupled receptor present a large family of integral membrane proteins that help transmitting signals into the cell[2]. They are activated by a set of highly chemically heterogeneous moieties, such as light, small organic molecules such as adrenaline, or peptides, lipids to name just a few. Structurally, GPCRs form a seven-transmembrane helical bundle joined by three intracellular and extracellular loops, an extracellular N-terminal domain of highly variable size and structure and a mostly flexible C terminus, that sometimes often contains an eighth helix[3, 4, 5]. Upon binding of an agonist a structural transition occurs through which the otherwise occluded binding surface for the G-protein becomes accessible[6]. While the importance of the pharmacology of GPCRs has been recognized in the 40ies by Ahlquist[7], the first high-resolution structure of a GPCR, of bovine rhodopsin, was published only in 2000[5], followed by the first X-ray structure of recombinantly produced GPCRs in 2007[8]. Since then we have witnessed an ever-accelerating avalanche of new GPCR structures, and the importance of these was acknowledged by the Nobel price in 2012 awarded to Levkowitz and Kobilka. In the meantime, even the structure of an agonist-bound GPCR coupled to a G-protein has been released[9], and attempts at solving the structure of a GPCR-arrestin complex are well underway.

While our knowledge of the structure of GPCRs in various states and their mode of activation and desensitization is rapidly increasing, detailed information on their folding pathway is still lacking. The refined three-stage model from Popot and Engelman postulates that secondary structure forms when the peptide chain partitions into the membrane-water interface[10-12]. Individual helices will then insert into the membrane, diffuses laterally within the membrane until native contacts are formed and the bundle fully assembles. Of course this biophysical model does not take into account participation of membrane-protein folding by the translocon machinery. Therein, the signal-recognition particle targets the nascent chains emerging from the ribosome tunnel to the translocon (Sec61p) complex[13]. In the co-translational folding model, segments of sufficient hydrophobicity are laterally gated from the translocon into the membrane interior, while formation of the helices may occur already within the translocon, or may even be initiated in the exit tunnel of the ribosome. Interestingly, a remarkable agreement of the biophysical Wimley-White scale[14] and the translocon scale from the van Heijne group[15] have been observed, indicating that formation of the TM helices and their insertion into the membrane are largely governed by hydrophobicity of the amino acids and hence by general properties amino acid sequence[16, 17]. Folding of polytopic membrane proteins can be described by a alternating series of events of helix-insertion into the

hydrophobic core and sequestering of loop-sequences into cytosolic or extracellular space[15]. What has become apparent is that this simple model is not always true but that helices may change their location during synthesis of later portions of the polypeptide chain[18-20].

In membrane proteins with helical bundles interhelical contacts are often formed between polar or even charged residues[21-23]. We noticed that some TM segments do not display favorable energies for full membrane insertion[24, 25]. In nascent chains in absence of some of the other helices these polar or charged residues would be exposed to lipids, and it may be questionable whether the TM parts then still insert into the membrane. We here therefore investigate conformational preferences of N-terminal fragments of a yeast G-protein coupled receptor (GPCR), the Ste2p receptor. The overall topology of GPCRs is fairly well conserved, and in particular the TM helices usually form the same contacts. Considering that proteins are synthesized starting with the N-terminus of the polypeptide chain, TM1 is expected to be the first segment to be inserted into the hydrophobic core, followed by TM2 and so on. We report on studies of polypeptides corresponding to the overlapping fragments TM1, TM12 and TM123. In the seven-helical bundle of GPCRs TM1, TM2 and TM7 form a core, and often more contacts exists between TM7 and TM1 and TM2 than between TM3 and TM2[4]. To investigate how TM7 influences helix stability and integration into the membrane, we have also looked at TM127, in which the N-terminal end of TM7 was directly linked to TM2. We determined the structures of these proteins by solution NMR methods. The investigated proteins contain overlapping stretches (e.g. TM1 is contained in all of them). We have additionally probed whether chemical shift assignments can be transferred from the smaller to the larger fragments. Considering that NMR spectra of membrane proteins of increasing size rapidly deteriorate in quality adapting chemical shifts from smaller fragments may present an interesting and useful technical advance in assigning the technically challenging membrane proteins.

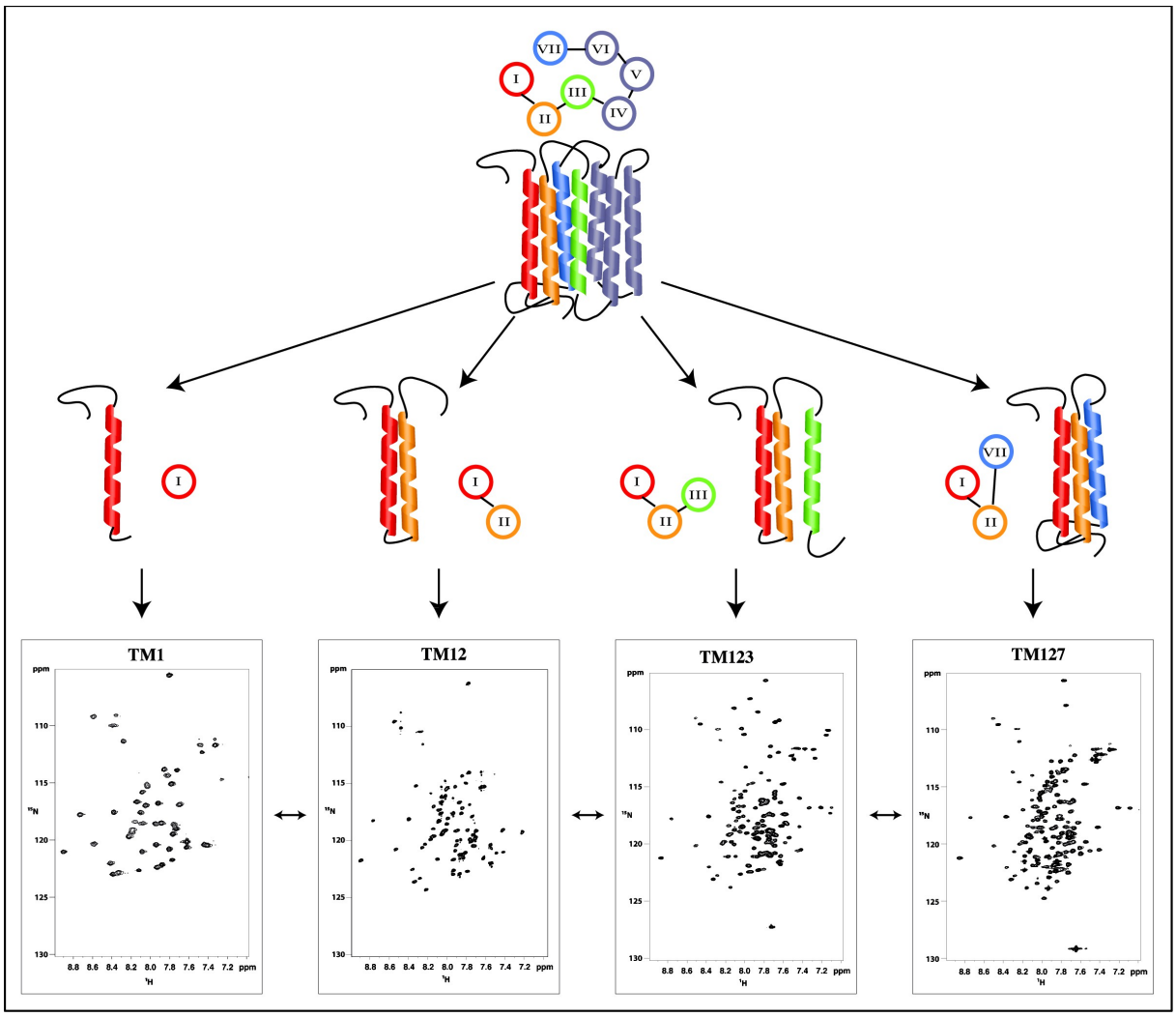


Figure 3.1: Schematic representation of the study of receptor fragments

The N-terminal half of the entire Ste2p receptor, schematically depicted on the top, is dissected into various fragments of increasing length. Typical [^{15}N,^{1}H]-HSQC spectra are depicted on the bottom.

3.3 Results

3.3.1 Protein expression and purification

Heterologous expression of membrane protein fragments often presents a significant problem due to their hydrophobicity and flexibility, which can result in misfolding and aggregation, and possible toxicity, which leads to overall low expression levels. Therefore, fusion proteins are often utilized to enhance over-expression and act as signaling sequences to direct the protein into specific cell compartments[26-30]. In the past our laboratory has mainly utilized two different fusions: The TrpΔLE sequence introduced by the Naider lab, as well as the N-terminal domain of the Y4 GPCR.

Whereas the former is highly hydrophobic, the latter presents a small (40 amino acid) soluble protein. Despite their different solubilities both fusions direct the expression product into inclusion bodies. Both fusion partners have to be removed after expression, and this task can be achieved using chemical or enzymatic cleavage. Chemical cleavage works best with cyanogen bromide, which however precludes occurrence of internal Met residues. Enzymatic cleavage is generally problematic since the fusion protein is highly hydrophobic and therefore must be solubilized in rather high concentrations of denaturant – and conditions must be chosen carefully to be compatible with retaining enzymatic activity of the protease. When starting this project with TM123 and TM127 TrpΔLE fusions were used as constructs, and the sequence contained only a truncated part of the N-terminal domain to reduce the overall size of the proteins. Problems with producing the single Cys mutants required for PRE measurements, due to side-reactions of the CNBr reagent with Cys residues, then forced us to investigate whether these proteins cannot be expressed directly, and this was indeed the case when adding the entire N-terminal domain to the constructs.

The Ste2p fragments TM1 (G31-T78), TM12 (G31-T110), TM123 (G31-R161) and TM127 (G31-T114, T274-L340) were expressed as C-terminal fusions to the TrpΔLE sequence. Cysteine and methionine residues have been replaced by amino acids with similar properties (Ser for Cys; Ile, Leu and Val for Met residues) to retain feasibility of chemical cleavage and preserve functionality of the receptor as reported before[31, 32] . Several *E. coli* expression strains (BL21 (DE3), BL21-AI, BL21 (DE3) pLysS, BL21 (DE3) STAR pLysS, BL21 Rosetta pLysS, C41 and C43) were screened for expression, as expression levels differed considerably between different fragments and strains, to achieve acceptable expression yields in minimal media. Inclusion bodies were solubilized in 70% TFA and, following chemical cleavage of the fusion proteins with CNBr, proteins of interest were purified using RP-HPLC. This resulted in expression yields of about 15mg and 5mg of purified protein per liter of $^{15}N^{13}C$-labeled and $^{15}N^{13}C^{2}H$-labeled M9 culture, respectively, for TM1 (using the strain BL21-AI), 11mg/L and 5mg/L for TM123 (BL21-AI)[33] and 15mg/L and 5mg/L for TM127 (BL21(DE3)STARpLysS). Expression details of TM12 have been reported previously[1].

To obtain additional long-range distance restraints for TM127, single cysteine mutants to specifically attach the paramagnetic spinlabel MTSL were also designed. As described in [34]about one spinlabel per TM helix is required to unambiguously establish the relative TM arrangements. Accordingly we have placed single-Cys residues at positions S47, S75 and S104. Single Cys mutants of TM127 were expressed directly containing the full N-terminus (S2-T114, T274-L340) as well as an N-terminal His_{10}-tag in BL21(DE3) cells. After RP-HPLC purification, 40mg of protein

per liter of ^{15}N-M9 medium culture were obtained. After coupling of the MTSL maleimide, the protein was again purified by Ni-NTA chromatography with a relative yield of 60%. For all proteins purity was checked by SDS-page and ESI-MS.

For NMR measurements, samples were measured in 150mM LPPG/DPC (4:1) micelles as the membrane mimetic, using 40mM potassium phosphate buffer at pH 6.4.

3.3.2 Backbone and side chain assignments

The NMR samples of TM1, TM123 and TM127 all exhibit good signal dispersion, when considering their highly alpha-helical nature, as well as homogeneous line-widths (see Figure 3.1 and Supp. Mat. Figures S3.11-14). Analogous residues in different fragments exhibit similar chemical shifts, which greatly facilitated assignment of the larger fragments (*vide infra*). Almost complete backbone assignment and partial sidechain assignment could be achieved for all fragments (*vide infra*).

In agreement with common experience we noticed that residues located in transmembrane regions generally exhibited dramatically lower intensities than their terminal counterparts, which, in peak clusters made it difficult to uniquely identify the number of peaks, as well as their correct position. In those cases, HNCO experiments proved to be critical to specify the correct number of peaks and their distinct positions. Further assignment problems occurred in loop regions and around charged/polar residues located in transmembrane helices, where conformational exchange is believed to cause line broadening. In general, no assignments were obtained for aromatic moieties.

TM1 was assigned using standard triple-resonance spectra such as the related HN(CO)CACB and HNCACB or HNCO and HN(CA)CO experiments[35, 36]. Sidechains were annotated by a combination of HCCH-TOCSY[37, 38] and ^{13}C-resolved NOESY spectra. Almost complete backbone (97%) and partial sidechain (66%) assignment could be achieved for TM1.

Assignment procedures for TM1-TM2 as well as details of its structure calculation have been published previously[1]. Samples containing TM123 exhibited a strong disposition to form soluble aggregates, with all TM123 samples in general being less stable compared to TM1 or TM127 samples. Depending on the labeling scheme of the protein as well as detergents, TM123 aggregated completely in a matter of hours, for the sample with reverse ILV methyl labeling in deuterated detergents, or up to several days, for samples without deuteration, which was confirmed by SEC-

MALS experiments (see Supplementary Material Figure S3.15). Nevertheless, we could assign about 95% of the backbone resonances of TM123, with the only missing residues being the two N-terminal amino-acids, as well as two prolines and two residues at the beginning of TM2 (N54 and Q55). In contrast, only 55% of sidechains could be assigned, the reason again being high residue redundancy and subsequent spectral overlap, as well as broad line-width resulting in low-quality HCCH-TOCSY and ^{13}C NOESY spectra.

In case of TM127, it was possible to achieve almost complete backbone assignment (93%), with the exception of the two N-terminal amino acids, as well as a region in the center of TM7 containing the Leu-Pro-Leu sequence. We suspect that this broadening is due to conformational exchange.. For TM127, it was possible to assign 63% of sidechain resonances. Most remaining problems during assignment were related to the high occurrence of certain amino acids, resulting in significant resonance overlap, especially for hydrophobic residues. Moreover, non-protonated sites in the otherwise perdeuterated detergents d_{36}-LPPG and d_{38}-DPC result in strong detergent signals from the central glycerol unit predominantly in the Cα region, thus further complicating assignments.

Through selective reverse methyl labeling of Ile, Leu and Val, most of the methyl groups could be assigned successfully, even though the remaining sidechain atoms could not be uniquely assigned. However, as is often the case with membrane proteins, only a small number of these assignments have yielded unambiguous NOE restraints.

3.3.3 Assignment by shift adaptation from fragments

TM12 and TM127 or TM123 constitute two pairs of proteins in which nearly 80 residues are identical, namely those from TM12. We have investigated whether chemical shift assignments can be obtained by adapting peaks from the previously assigned, shorter fragment TM12 to the longer one. Corresponding peak pairs in the two [^{15}N,^{1}H]-HSQC spectra are identified based on two criteria: First, we assume that the peak in the spectrum of the longer construct, e.g. TM127, will be close to the position of the peak in the shorter TM12 fragment. However, there will still be a number of peaks in vicinity to the original peak that may present the correct solution. The second criterion therefore comprises a peak pattern match of the strip in the 3D ^{15}N-resolved NOESY corresponding the original peak anchor with those from potential candidates in the TM127 spectrum. Amide moieties will be close to protons from its own sidechain or from neighboring

residues, but, taking the larger distance to sidechain protons from *other* helices into account, likely not from sidechains that are close due to tertiary contacts. We suspected that chemical shifts of sidechains will not change substantially unless secondary structure is formed or changes. In that respect it is important that our previous work demonstrated that secondary structure in fragments is largely correct. The procedure is schematically shown in Figure 3.2:

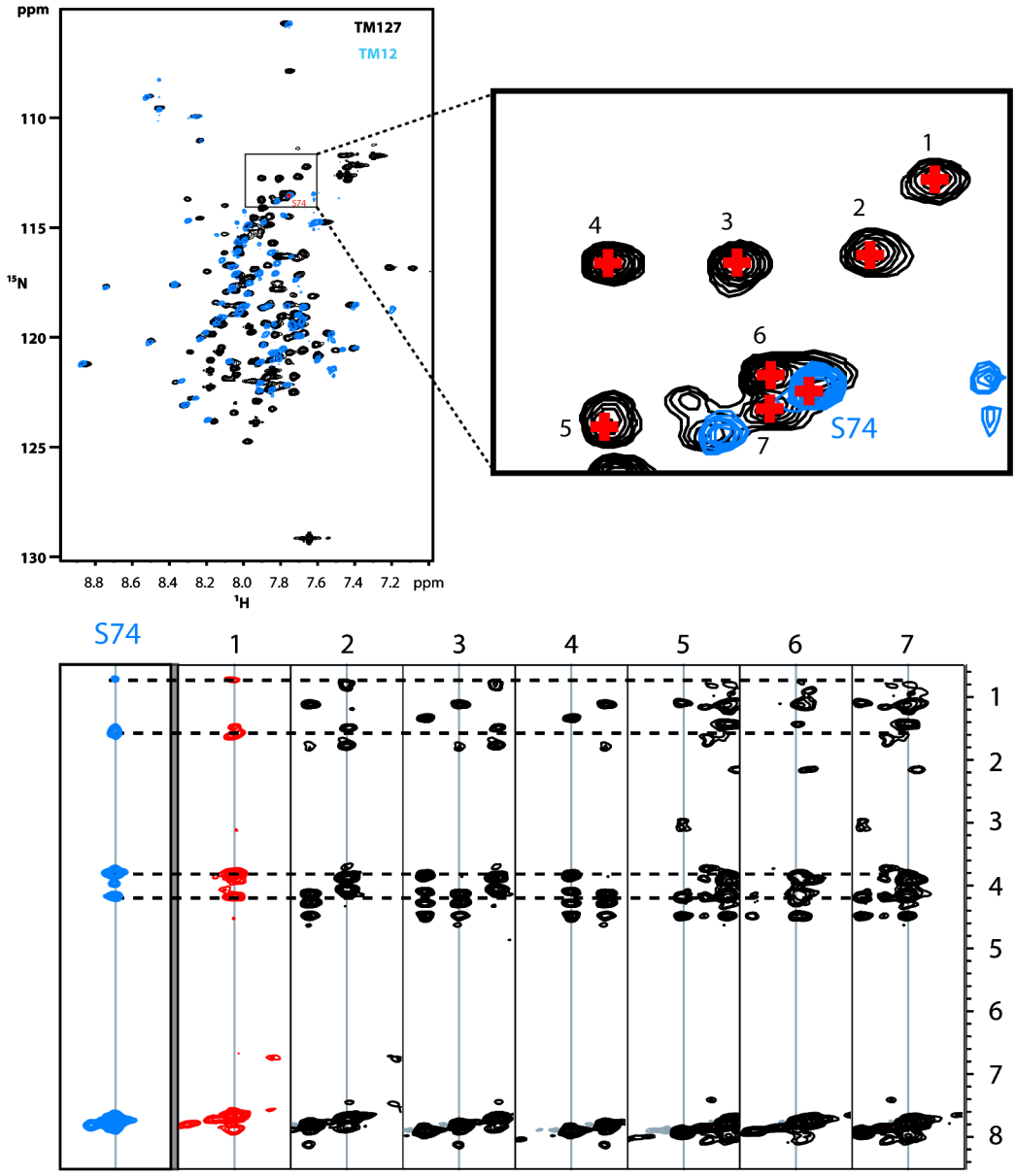


Figure 3.2: Representation of automatic strip matching

Overlay of the [^{15}N,^{1}H]-HSQC spectra of TM12 and TM127 (top). 3-D ^{15}N-NOESY strip comparison (bottom). 3-D NOESY strips from assigned peaks in TM12 (blue) are compared to possible candidates of TM127 (black)

In our initial trials using only spectra from the 700 MHz spectrometer and ^{15}N NOESY spectra with only 22 complex points in the ^{15}N dimension only approximately 65 % of amide peaks were correctly adapted. The major problem originated in peak overlap in the [^{15}N,^{1}H]-HSQC spectrum. Due to the insufficient resolution in the ^{15}N domain NOE correlations could not be unambiguously attributed to individual peaks, and hence much cross-talk (missing or wrong NOE correlations) appeared in the strips. When [^{15}N,^{1}H]-HSQC and 3D ^{15}N NOESY spectra were recorded at 900 MHz, this time using 192 complex points in the ^{15}N domain 70 (93 %) out of the 75 strips of TM12 have been correctly matched to the total of 144 strips of TM127 (see Figure 3.3). We like to emphasize here that both [^{15}N,^{1}H]-HSQC spectra as well ^{15}N-NOESY spectra were automatically picked and integrated.

Unsuccessful cases occurred when peaks in either the original TM12 or target [^{15}N,^{1}H]-HSQC spectra overlapped. In these cases strips will contain NOEs originating from two or perhaps even more amide protons. Whether a correct assignment solution is then found depends on how different the hypothetic strips are, and hence how much they are deteriorated by adding "wrong" NOEs. Other misassignment occur when the location of the target peak is either very different, as encountered by residues at the termini, for which chemical shift are rather different, or for peaks in highly overlapping (central) parts of the [^{15}N,^{1}H]-HSQC spectrum.

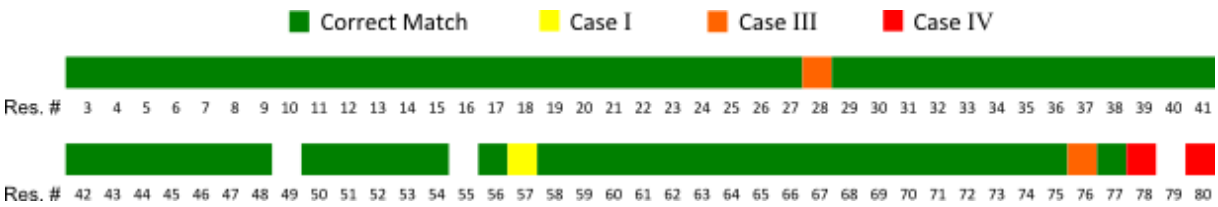


Figure 3.3: Success of 3D NOESY strip matching of TM12/TM127

Green and yellow coded residues are matched to the correct position, while orange and red colored residues are matched incorrectly. For details see text. Uncolored residues are not assigned in TM12

3.3.4 Conformational preferences

Usually, structure calculations of helical membrane proteins suffer from an insufficient number of long-range restraints due to the fact that complete sidechain assignments are difficult to obtain. More importantly, sub-optimal packing of helices, deficiencies of the detergents to perfectly mimic biological membranes and the inherent flexibility result in exchange broadening that tends to damp out the weak but structurally important long-range NOEs. To compensate for that RDCs,

PREs as well as chemical shift restraints are applied. Since the backbone assignment of TM127 was rather complete predictions from TALOS-N resulted in a fairly high number of backbone torsional angle restraints. Since those restraints only define secondary structure we complemented them with PRE data from two sets of spinlabel derived distance restraints originating from spin labels attached a the end of TM1 and at the end of TM2. In addition, NH dipolar couplings were measured to learn about the relative helix orientations. We have also probed solved access from a water-soluble spinlabel. All these data were used to obtain restrains for the final structure calculations. Back-prediction of the raw data from conformers of the NMR ensemble allowed judging how well primary NMR data were represented by the ensemble. In case of TM1 the structure calculations used NOEs-derived distance restraints and chemical–shift derived torsion angle restraints. In case of TM123 we have only probed secondary structure from NOEs and chemical shifts.

3.3.5 Chemical shifts

The propensity to form helices predicted from backbone chemical shifts is depicted in Figure 3.4. In general, all putative helices seem to be in place. However, it can clearly be seen that TM1 is destabilized in the center of the helix around the G-XXX-G motif, which becomes rigidified upon packing against TM2 in TM12. The N-terminal end of TM2 also seems to be destabilized in TM127 as well as in TM123, where some peaks are missing, possibly indicating the presence of conformational exchange. A short surface-associated helix at the N-terminus encompassing the segment from Phe-38 to Val-45 is also visible in all fragments. The loop regions between TM2 and the third helix in both TM123 and TM127 show some disposition to form helical secondary structure. In TM127, secondary structure in the center of TM7 is not well defined, largely due to the missing assignments for residues Leu-Pro-Leu. The fact that no peaks were found for Leu-289 and Leu-291 indicates that conformational transitions occur close to the central Pro residue. Similarly, the center of the third helix in TM123 is destabilized around several polar residues.

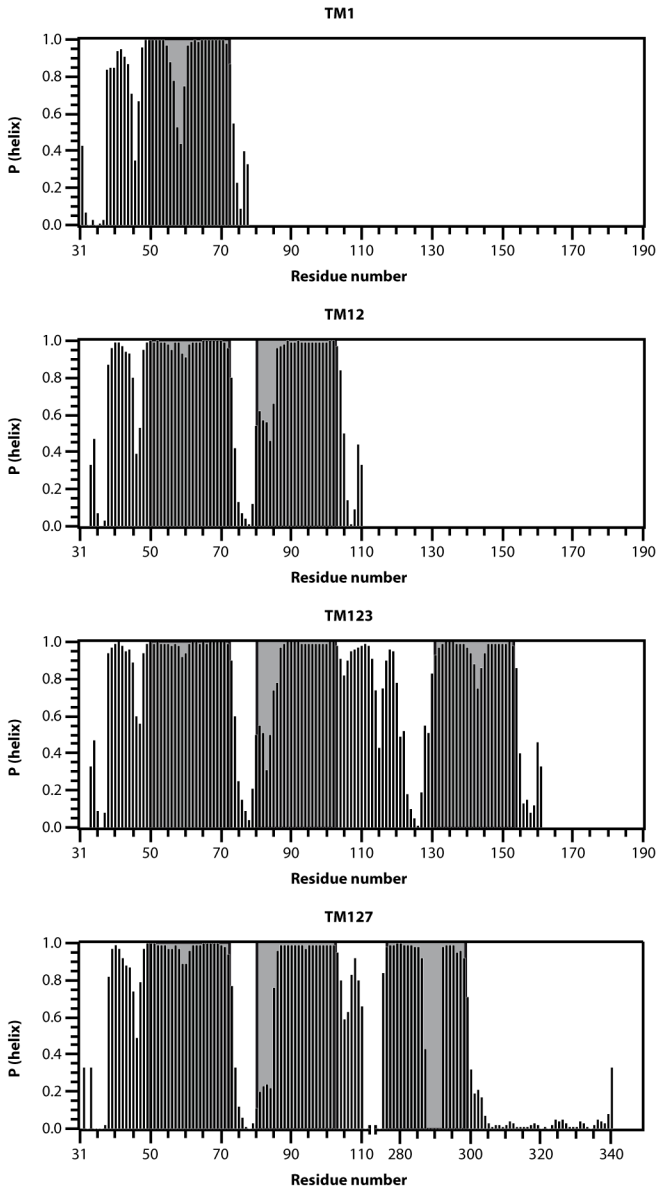


Figure 3.4: Secondary structure predictions obtained from TALOS

Secondary structure predictions obtained from TALOS for TM1, TM12, TM123 and TM127. The propensity of residues for forming α-helices is shown

3.3.6 Paramagnetic relaxation enhancements

To obtain long-range distance restraints derived from PREs, *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL), was attached to unique Cys sites in the protein. Unfortunately, the presence of cysteine residues was incompatible with successive chemical CNBr cleavage, and therefore His-N-TM127, a construct containing the full native N-terminus of the Ste2p receptor, that could be directly expressed (His_{10}, S2-T114, T274-

L340), was used instead. We designed three different single-cysteine mutants of this construct: S47C at the N-terminus of TM1, S75C in the short loop between TM1 and TM2, and S104C at the beginning of the relatively long loop between TM2 and TM7. Signals from residues close in sequence to the spinlabel completely disappeared in two of three cases, indicating quantitative coupling and full functionality of the spinlabel for constructs S74C and S104C. For some residues, signal intensities in PRE samples were as large as 120%, suggesting an error of up to 20%. Therefore, all PREs were classified in three different classes: Strong PREs, with less than 10% of the original signal intensity (<14Å), medium PREs, with residual peak intensities between 10-80% (14-20Å) and unaffected residues, with residual intensities larger than 80% (>20Å) following methodology originally introduced by the Wagner lab[39].

The S47C spinlabel resulted in incomplete bleaching of resonances from residues close in sequence, indicating that the spinlabel was partially deactivated. Therefore, no restraints derived from this spinlabel were used for structure calculations. The PREs from the S75C and S104C mutants are summarized in Figure 3.5. The S75C mutant exhibited decreasing attenuation throughout the whole TM2 helix and the first half of TM1. Signal attenuations can also be observed at the end of TM7 and the C-terminus, revealing spatial proximity between the first intracellular loop and the C-terminal end of TM7. The S104C single cysteine mutant also display linear attenuations extending towards the ends of TM2 and TM7, while showing weak PRE effects to the N-terminus of TM1, as well as medium attenuations for residues of the surface associated short N-terminal helix, indicating weak interactions between the second loop and the N-terminus.

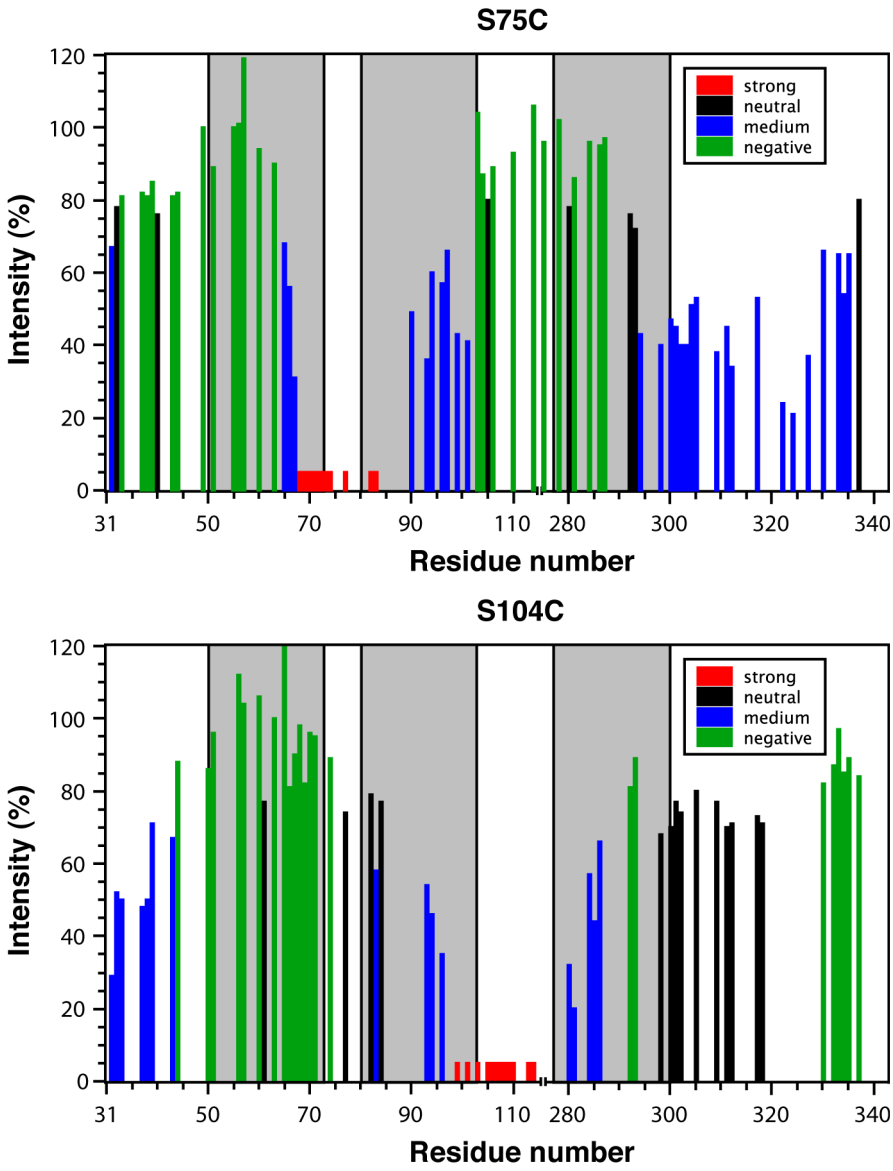


Figure 3.5: Residual intensities for site-directed spinlabels

Residual intensities for site-directed spinlabels S75C (top) and S104C (bottom). Red bars indicate residues with no observable intensities; blue bars residues up to 80% and green bars residues above 80% residual intensity.

3.3.7 Probing relative helix orientations by RDCs

Residual dipolar coupling (RDC) measurements present a valuable tool to determine the vector orientation between two nuclei, relative to a molecular axis system, leading to global structure information. In this case, dinucleotide 2'-deoxyguanylyl-(3',5')-2'-deoxyguanosine (dGpG), which forms G-tetrades in the presence of potassium, was used as the alignment agent. RDCs for the backbone amide moieties of TM127 were measured using the TROSY-based ARTSY

experiment, and resulted in RDC values between -24Hz to +14Hz (Figure 3.6). TM1 and the N-terminal part of TM7 display negative RDCs throughout, while those from TM2 are positive, indicating that TM1 and TM7 are largely collinear and TM2 is roughly orthogonal to these. Frequent sign changes in the C-terminal half of TM7 indicate sampling of multiple conformations. Negative RDCs can also be observed for the short N-terminal helix and the second loop. RDCs from the mostly unstructured C-terminus have been omitted from the further analysis.

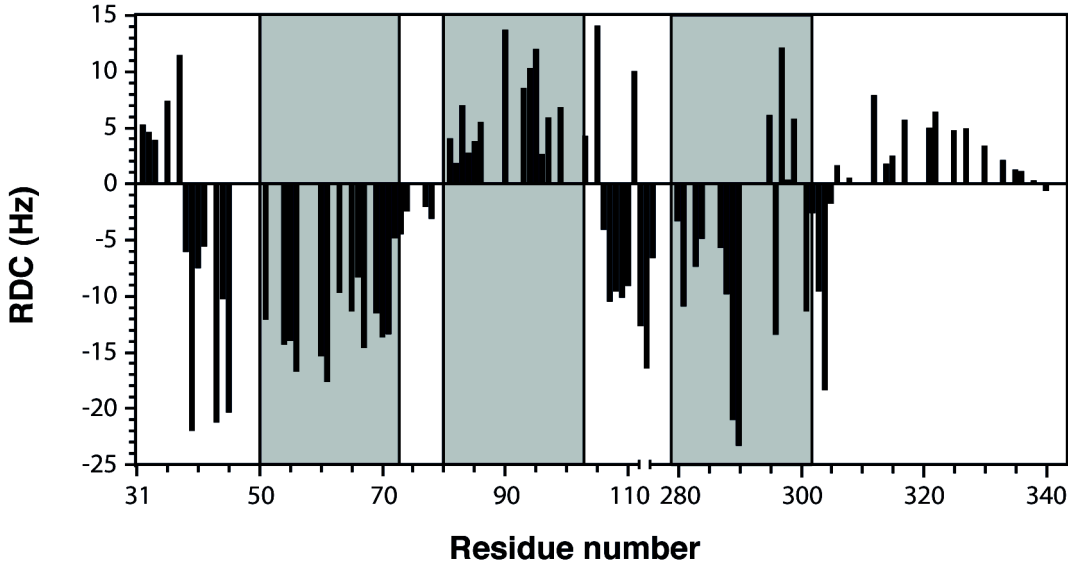


Figure 3.6: RDCs for TM127

Intensity of H-N RDCs for TM127 in Hertz, cataloged by residue.

3.3.8 Probing the integration into the micelle from water-soluble spinlabels

To map the topology of membrane-integration of TM123 and TM127, the soluble, inert paramagnetic probe Gd-(DTPA-BMA) (Gadolinium diethylenetriamine-pentaacetic acid bis-methyl amine) was added to the NMR-sample. In general, residues outside the micelle or closer to the micelle surface experience stronger signal attenuations compared to residues that are buried in the micelle interior. To minimize integration errors that result from peak overlap, PREs were quantified using HNCO experiments. At 15mM Gd-(DTPA-BMA) water exposed residues exhibit less than 20% residual signal intensity, while residues that are buried in the micelle still display intensities of up to 90%.

Overall, the transmembrane regions of both fragments, TM123 and TM127, seem to be well-integrated into the micelle, while the termini as well as the loop regions are located outside of the micelle. Signals from the short N-terminal helix as well as part of the second loop display moderately weaker attenuations, indicating that they interact with the micelle surface. TM1 appears to be the firmly embedded, especially in TM127, where most signals display minimal signal attenuations while TM2 is slightly destabilized relative to TM1, especially at the termini. TM7 in the TM127 fragment displays limited signal attenuation at the N-terminus, with residual peak intensities decreasing towards the C-terminus.

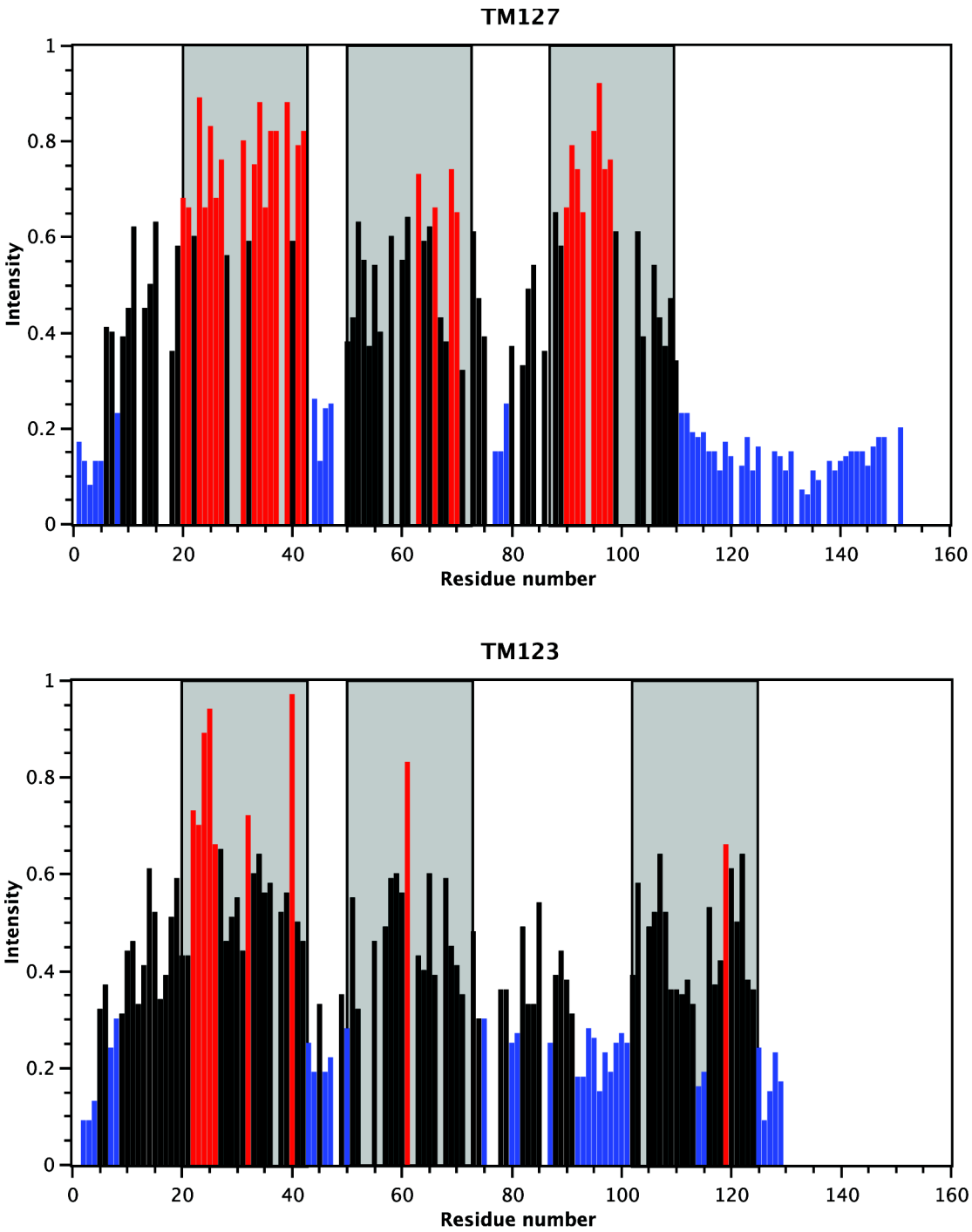


Figure 3.7: Residual intensities using soluble spinlabels

Relative HNCO peak intensities of TM127 and TM123 after addition of 15mM Gd-(DTPA-BMA). Red bars indicate residual intensities above 65%, blue bars those under 30%

In case of TM127 restraints for structure calculations were initially derived from these PREs, but omitted for later structure calculation rounds, as they seemed to over-restrain the helix bundle. Residues with signals weaker than 30% (blue), which are water exposed, were defined with a lower

limit of 19Å to the micelle center, and residues with signals stronger than 65% (red), which are integrated into the micelle, were defined with an upper limit of 15Å.

The Gd-(DPTA-BMA) data indicate that TM123 generally seems to be less well-folded compared to TM127. For example, the center of TM3 is partially solvent exposed, in agreement with the occurrence of polar residues in that part. The poor packing of the helices in TM123 may explain the instability and proclivity of TM123 to aggregate.

3.3.9 Structure calculation

The structure determination for TM127 did not yield one unique structure, indicating that TM127 is a highly dynamic system, not a rigid helix bundle, which is clearly supported by the PRE measurements. Whereas the helices appear to be firmly integrated into the micelle (as can be seen above in Chapter 3.3.8), interhelical contacts seem to be transient. The observation of mirror topologies (those with a wrong sense of relative helix arrangements) also implies that TM2 and TM7 compete for the same interhelical contacts in TM1. In general, it can be seen that TM2 packs against TM1 at an angle, which is confirmed by the RDC data, whereas TM7 occupies a multitude of conformations, mostly exhibiting a kink around the Leu-Pro-Leu sequence.

To obtain a representative statistical analysis of the different conformations of TM127, we calculated 5000 structures, clustering the 500 structures with the lowest energies based on their backbone RMSD (<3Å) in the helical regions. This resulted 48 clusters contained up to 95 structures. 25 clusters accommodated three or less structures, and were therefore too lowly populated and hence omitted from further analysis. As mentioned above, mirror topologies compared to the 'correct' topology derived from a homology model can be observed in roughly 25% of the structures. The four most prominent clusters with at least 30 structures have been analyzed by back-calculation of the raw PRE and RDC data and subsequently compared to the experimental values, as well as to the back-calculated values for the structure based on a homology model (Figure 3.8). Whereas the simulated RDC values of all four clusters only exhibited differences of up to 3Hz compared to the experimental data, RDCs based on the model deviated up to 18Hz (see Supplementary Material Figure S3.16).

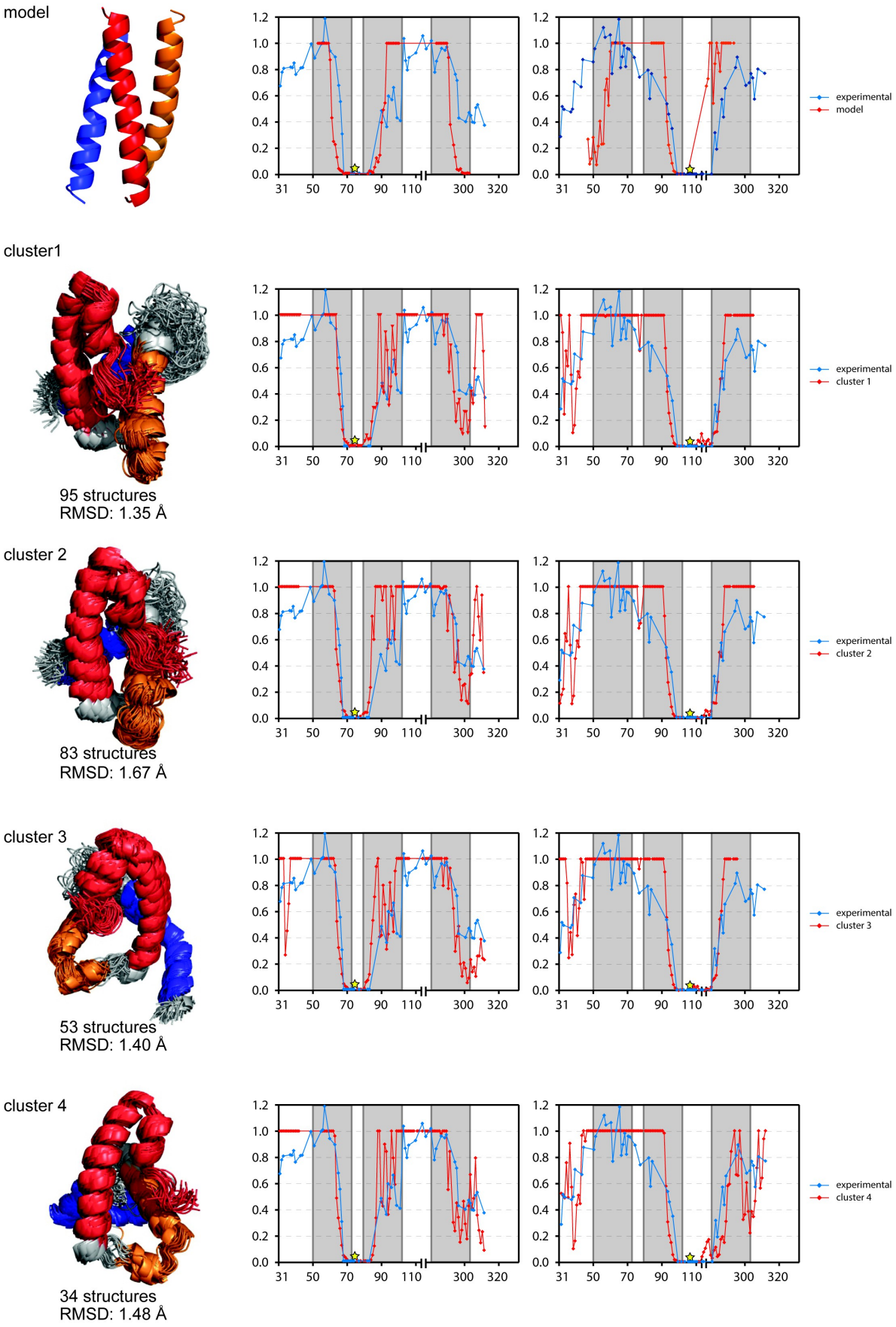


Figure 3.8: PRE simulations

PRE simulation for the 4 largest clusters from the structure calculation as well as for the homology model. Structures containing TM1 (red), TM2 (orange) and TM7 (blue) are shown on the left. The back-calculated PREs are shown in red for the S75C spinlabel (center) and for the S104C spinlabel (right). The respective experimental values (blue) can be found in each diagram

In comparison to the simulated values derived from the homology model, the experimental restraints show much gentler slopes, especially from the S75C spinlabel affecting TM2, which indicates dynamic averaging from different confirmations. Our data therefore indicate, that TM127 in micelles does not exist as a rigid helix bundle. It also has to be mentioned that the simulated PREs from the four largest clusters of the structure calculations indicate proximity of the end of TM7 and the S75C spinlabel, as signals at the C-terminus of TM7 are almost completely attenuated (less than 20% residual signal). This does not correspond to the experimental data, where the equivalent signals still exhibit over 40% of the original signal. The reason for this can likely be found in conformational averaging, as the C-terminus of TM7 does exhibit conformational exchange around the Leu-Pro-Leu region, and hence may not pack against TM1 at all times.

One problem with the structure calculation of dynamic systems is the fact that restraints, which are derived from transient data, are averaged. In a structure calculation they are treated to fit for a single, unique structure and not to an ensemble of different conformers. However, in systems, where one part of the protein is flexible, as is expected for the C-terminal part of TM7 in this case, it is much more likely that a restraint is only fulfilled for the population-weighted average of the ensemble. To counter this problem we are currently trying to establish multi-state structure calculations. In these calculations, the sequence is multiplied several times and restraints only have to be fulfilled in at least one of those structures.

3.4 Materials and methods

3.4.1 Chemicals and solutions

$^{15}NH_4Cl$, ^{13}C-d_7-D-Glucose and D_2O were purchased from Spectra Stable Isotopes (Andover, Massachusetts, USA). 1-palmitoyl-2- hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LPPG)/ dodecyl-phosphocholine (DPC) were bought from Avanti Polar Lipids (Alabaster, Alabama, USA). All other chemicals were obtained from Sigma-Aldrich (Buchs, Switzerland).

3.4.2 Cloning procedures:

Construction of the Ste2p TM127 (*G31-T114, T274-L340*) expression vector was accomplished using a two-step process. PCR was performed using primers designed to remove residues G115-G273 and the pBluescriptSK(-) BEC2(M54L, M69V, M71I, M165I) Ste2p vector as a template. The resulting sequence was PCR amplified using a forward primer containing a *Hind*III site and a sequence complementary to the N-terminal sequence of the TM127 construct, and a reverse primer containing a *Bam*HI site and a sequence complementary to the C-terminal sequence of the TM127 construct. The amplified fragment and the pSW02 vector were digested with *Hind*III and *Bam*HI, gel purified, ligated using T4 DNA ligase, and transformed into DH5α. DNA from some of the resulting colonies was purified and sequenced. The resulting plasmid, pSW127, contained the TrpΔLE leader sequence in phase with Ste2p(G31-T114,T274-L340, M54L, C59S, M69V, M71I, M294L).

3.4.2.1 Cloning of TM127 cysteine mutants for PRE labeling

To test direct expression of TM127 three different expression vectors were constructed: One containing the full native N-terminus of Ste2p (S2-T114, T274-L340), as well as two constructs additionally containing either a N- and C-terminal His_{10}-tag. The N-TM127 fragment was derived from the expression vectors pSW127 and pRE01. The latter plasmid contains the N-terminus in front of the TM1-TM3 fragment. Both plasmids were used as templates to construct N-TM1-TM2-TM7 using suitable overlapping primers in the TM1 region, a forward primer containing a *Nde*I restriction site and a reverse primer containing a *Bam*HI site. The amplified fragment as well as the pSW02 vectors were digested with *Nde*I and *Bam*HI, gel purified, ligated using T4 DNA ligase, and transformed into DH5α cells. Several colonies were picked and sequenced, resulting in the verified

pMP01 expression vector containing N-TM127. N- and C-terminal His_{10}-tags were introduced using FX cloning[40]. The N-TM127 insert was amplified using the pMP01 vector as template with primers encompassing *Sap*I restriction sites. Precompiled FX cloning vectors were used, containing ORF-verified sequences of either N- or C-terminal His_{10}-tags with a C3 protease cleavage site adjacent to the HIS_{10}-tag. The insert and vectors were mixed, digested using *Sap*I, ligated using T4 DNA ligase, transformed into DH5α cells and sequenced, resulting in expression vectors pMP02 (His-N-TM127) and pMP03 (N-TM127-His).Point mutations of single Cys residues were introduced using standard Quickchange® protocols. Using the vector containing the HIS-N-TM127 insert as a template, three cysteine mutants were designed, His_{10}-N-TM127(S47C), His_{10}-N-TM127(S75C) and His_{10}-N-TM127(S104C).

3.4.2.2 Cloning of TM1

The Ste2p TM1(G31-T78) construct was cloned as a direct expression vector from a vector encoding TM1-TM3, that itself was constructed from the vector pKC01, that contains the ΔTrp fusion of TM1-TM3. The TM1-TM3 direct expression vector, pRE01, was constructed by PCR amplification of the G31-R161 region of the Ste2p sequence from the pKC01 vector using a forward primer containing a His tag and an *Nde*I restriction site and a TM3 reverse primer containing a *Bam*HI restriction site. The resulting PCR product and the pKC01 vector were digested with *Nde*I and *Bam*HI. The digests were gel purified and then ligated using T4 DNA ligase. The resulting reaction mixture was transformed into DH5α and the DNA was purified and sequenced. The amplified plasmid, pRE01, contained a 6 residue His tag in frame with Ste2p(G31–R161, M54L, C59S, M69V, M71I).

Double stranded mutagenesis was performed on this pRE01 plasmid using primers designed to incorporate two stop codons after residue T78. The resulting PCR product was digested with *Dpn*I. The digested DNA was transformed into DH5α, ten colonies were isolated, and the plasmid DNA was extracted from each of the colonies. The extracted DNA was sent for sequencing. The resulting expression vector, pKC03, contained a 6 residue His tag in frame with Ste2p(G31-T78, M54L, C59S, M69V, M71I).

3.4.3 Expression

3.4.3.1 Expression of TM123 and TM127

As expression in BL21-AI cells was found to be optimal for the fusion protein ΔTrp-TM1-TM3[33], initial expression attempts for ΔTrp-TM127 were also made in these cells, but a small-scale test expression revealed that the construct did not express well. Following further optimization the protein was finally expressed in BL21 Star(DE3)pLysS, using 0.5 mM IPTG as the inducer and induction at 22°C for 18h.

ΔTrp-TM1-TM3 was also expressed to selectively protonate the methyl carbons of the isoleucine, leucine and valine (I, L, V, respectively) in an otherwise perdeuterated background[41]. This was accomplished using 2-keto-3-methyl-d,3-d_1-1,2,3,4-^{13}C butyrate and 2-keto-3-d_2-1,2,3,4-^{13}C butyrate as the metabolic precursors for V and L and I, respectively. The pKC01 expression vector was transformed into BL-21 AI cells. A small pre-culture was prepared in LB and used to inoculate 100 mL of M9 minimal medium prepared with $^{15}NH_4Cl$, [^{13}C]-glucose, and D_2O. The 100 mL culture was grown to OD_{600} of 0.4 and then diluted to 200 mL with M9 minimal medium prepared with $^{15}NH_4Cl$, ^{13}C-glucose, and D_2O. The 200 mL culture was grown to OD_{600} of 0.4 and then diluted to 1L with M9 minimal medium prepared with $^{15}NH_4Cl$, [^{13}C]-glucose, D_2O, and the labeled I, L, V metabolic precursors. The cells were again grown to OD_{600} of 0.4, expression was induced with 0.2% L-arabinose, and the culture was incubated at 37°C for 9.5 h. After the incubation, the cells were pelleted by centrifugation, and IBs were prepared as previously described[42].

For selective reverse methyl labeling of TM127, the pSW127 plasmid was transformed into BL21 Star(DE3)pLysS cells. 1 mL of LB (H_2O)/Cam,Amp overnight culture was used to inoculate 6 mL of LB (D_2O)/Cam, Amp culture and grown overnight. 1.5 mL of this overnight culture was subsequently transferred to 40 mL M9 (D_2O)/50% Cam, 20% Amp medium with $^{15}NH_4Cl$ as well as [^{13}C]-glucose, grown to OD_{600} of 1.2 and added to 1L of M9 medium. After growth to an OD_{600} of 1.0, the culture was cooled to 22ºC, metabolic precursors were added, and expression was induced with 0.5 mM IPTG for 18 h. Subsequently, the cells were harvested, lysed by ultra-sonication for 10 minutes, inclusion bodies were isolated by centrifugation at 14 000 rpm for thirty minutes and aliquoted into 8 equal parts to not overload the column.

3.4.3.2 Direct Expression of TM1

Direct expression of TM1 was done in BL21-AI cells. Two 500 mL expressions in M9 medium containing $^{15}NH_4Cl$ and $^{15}NH_4Cl/^{13}C$-glucose, respectively, were conducted with induction at 37°C overnight. The entire culture volume for each expression was pelleted and inclusion bodies were prepared as described below.

3.4.4 Cleavage and Purification of TM1-TM3 and TM127, TM1

For the purification of TM1-TM3 and TM127, inclusion bodies of the ΔTrp-fusions were solubilized in 70% TFA and the TrpΔLE was removed chemically by incubation with 75mM CNBr for 90 minutes at room temperature. The mixture was then directly injected onto a Sorbax 300SB-C3 RP-HPLC column at 60ºC. The protein was purified, using a gradient from 40% to 100% buffer B (buffer A: H_2O/ 10% isopropanol/ 0.1% TFA; buffer B: acetonitrile/ 10% isopropanol/ 0.1% TFA) and subsequently lyophilized as described before (Caroccia 2011).

Purification of the TM1 from direct expression was performed using RP-HPLC with a water/acetonitrile/isopropanol gradient at 60°C. The reservoirs contained water/isopropanol/0.1% trifluoroacetic acid (TFA) and acetonitrile/ isopropanol/0.1% TFA and the gradient was from 36%–72% acetonitrile with a constant isopropanol concentration of 10%. Preparative HPLC was performed on a Zorbax 300SB- C3 Prep RP-HPLC column at 60°C using gradient elution with a water/acetonitrile/isopropanol/ 0.1% TFA gradient from 36%–81% acetonitrile with a constant 10% isopropanol concentration.

3.4.4.1 Direct Expression and Purification of TM127 cysteine mutants

To determine the best possible conditions for direct expression of TM127 fragments, three expression vectors, namely pMP01 (N-TM127), pMP02 (His_6-N-TM127) and pMP03 (N-TM127-His_6), were transformed into a number of different E.coli cells. 10 mL LB (H_2O) test expressions at 37ºC and 20ºC for each transformation were evaluated, from which it was concluded that only BL21 (DE3) cells exhibit reliable over-expression. Single cysteine mutants were expressed in 500 mL M9 (H_2O)/50% Kan medium with $^{15}NH_4Cl$, grown to an OD_{600} of approximately 0.7, cooled to 20ºC and induced with 0.5 mM IPTG. Cells were harvested after 16h and lysed using ultra-

sonication. Inclusion bodies were isolated by centrifugation at 14 000 rpm for 30 minutes, and aliquoted into six equal parts.

Each inclusion body aliquot was dissolved in 70% TFA and injected directly onto a Sorbax 300SB C3 RP-HPLC column. The protein was purified using a gradient from 40 – 100% buffer B (buffer A: H_2O/ 10% isopropanol/ 0.1% TFA; buffer B: acetonitrile/ 10% isopropanol/ 0.1% TFA) and subsequently lyophilized.

3.4.5 MTSL spinlabel coupling to single cysteine mutants

Coupling of the MTSL spinlabel to single cysteine mutants was achieved by first dissolving 0.4 mM protein in 40mM Tris/HCl buffer at pH 7.5 containing 6M Guanidinium-Hydrochloride (GdmCl) , followed by addition of 2mM DTT and gentle shaking for 4h at room temperature to assure complete reduction of cysteines. After the solution was purged with nitrogen to remove oxygen for several minutes, a 15- to 20-fold excess of MTSL spinlabel dissolved in DSMO was added, and the solution was incubated at room temperature over night.

The reaction mixture was injected directly onto a Ni-NTA column and washed with about 100 column volumes of the reaction buffer to remove unreacted spinlabel and residual DTT. The reaction product was eluted with 500mM imidazole after another wash step with 10mM imidazole, and collected fractions were dialyzed against water and lyophilized.

For NMR sample preparation the protein was dissolved in hexafluorisoropanol (HFIP)/H_2O (50:50), and half of the solution was deactivated with 10mM ascorbic acid and afterwards neutralized with sodium hydroxide. Subsequent sample preparation was carried out as described below.

To evaluate the effect of PREs, the ratio of peak intensities in [^{15}N,^{1}H]-HSQC spectra measured on freshly prepared samples was computed relative to samples in which the MTSL label was deactivated by ascorbic acid. Peaks that were not sufficiently separated and therefore could not be integrated reliably were omitted from further analysis.

3.4.6 NMR sample preparation

NMR samples contained 40 mM $K_3(PO_4)$ buffer at pH 6.4, and 120 mM LPPG/ 30 mM DPC and were produced using a protocol slightly modified from the one described by Killian[43]. Therein, the protein was first dissolved in small amounts of HFIP/water (8:1), while detergents were dissolved in a phosphate buffer equivalent to their final sample concentrations. After mixing the two solutions, water was added stepwise to dilute the organic solvent and to allow micelles to form. After two lyophilization steps, the sample was taken up in H_2O/D_2O (9:1) the final sample volume. In case of TM1-TM3, several cycles of dissolving the sample in HFIP and H_2O and subsequent lyophilization were necessary to achieve acceptable protein integration into the micelle.

To samples that were used to probe micelle integration, corresponding amounts from 100-times concentrated Gd-(DPTA-BMA) stock solution have been added. Gd-(DTPA-BMA) concentrations of 5mM, 10mM and 15mM were measured and compared to a blank of 1mM to reduce effects resulting from T1 relaxation.

3.4.7 NMR spectroscopy

All samples for assignment purposes were measured at 317K on a Bruker AV700 spectrometer equipped with a triple-resonance cryoprobe at 317K NMR samples contained 0.2 – 0.4 mM protein in 40mM phosphate buffer, pH=6.4 , 120 mM LPPG and 30 mM DPC as described previously[1]. ^{15}N NOESY spectra that were used for strip matching were recorded on an Avance-II 900 MHz spectrometer in the European NMR facility in Frankfurt. Proton chemical shifts were referenced to the water line at 4.47 ppm at 317 K, from which the nitrogen and carbon scales were derived indirectly by using the conversion factors of 0.10132900 (^{15}N) and 0.25144954 (^{13}C).

Sidechain assignments were obtained from standard TROSY-type triple-resonance experiments. Briefly, an HNCO data set was used to pick peaks in the [^{15}N,^{1}H]-TROSY spectrum in order to early on recognize peak overlap, and to adjust peak positions in cases of peak overlap. Backbone and partial sidechain assignments were then obtained from HNCACB and HN(CO)CACB experiments (or HNCA and HN(CO)CA) spectra. Sidechain resonance assignment was accomplished using hCCH-TOCSY/COSY[44, 45] in combination with [^{13}C,^{1}H]-HSQC and ^{13}C-resolved aliphatic/aromatic-NOESY experiments[46]. Spectra for sidechain assignments required the use of d_{36}-LPPG and d_{38}-DPC to eliminate the strong residual signals from detergent that would otherwise obscure the region of C-α and methyl resonances. We noticed small changes in peak

positions between spectra measured on deuterated and non-deuterated proteins as well as for deuterated and non-deuterated detergent. Therefore, Cα and Cβ chemical shifts obtained from the backbone assignments were initially adjusted by using hCCH-TOCSY or hCCH-COSY spectra. All chemical shifts were finally correlated to peak positions in the [^{15}N,^{1}H]- and [^{13}C,^{1}H]-HSQC spectra. Spectra were processed within the Bruker spectrometer software Topspin 2.1 and chemical shift assignments were performed using the software CARA[47].

Studies of membrane insertion topology were performed using the micelle integrating spin label 4-(3-ƛ^{1}-oxidanyl-2,2-dimethyl-4-tridecyl-1,3-oxazolidin-4-yl)butanoic acid (5-DSA) (Sigma). Samples for PRE measurements were prepared by addition of concentrated stock solutions of the paramagnetic substance to the micellar protein solution containing ^{15}N-labelled TM1. After addition the pH was readjusted to pH 6.0. The final concentrations of 5-DSA was 6mM.

3.4.7.1 RDC

RDC samples were obtained by dissolving dinucleotide 2'-deoxyguanylyl-(3',5')-2'-deoxyguanosine (dGpG) in 40mM K_3PO_4 buffer (pH 6.4), 150mM LPPG:DPC (4:1) containing 0.4mM ^{15}N^{13}C^{2}H-HIS-N-TM127 and 100mM KCl to ensure G-tetrade formation, to a final concentration of 30mg/ml. Several heating-cooling (45º-4ºC) cycles were necessary to completely dissolve dGpG. Formation of a stable liquid crystal phase was determined by observing the residual $^{2}H_2O$ quadrupol coupling on the solvent signal (see Supplemental Material S3.17).

The RDC tensor was determined using Singular Value Decomposition (SVD), which is implemented in the program CYANA. Back calculation of RDCs with the program MODULE[48] showed good agreement with the experimental values (see Supplemental Material Figure S3.18).

3.4.7.2 Assignment via adaptation of chemical shifts from overlapping fragments

The resonance signals of the 3D ^{15}N-NOESY spectra were picked and integrated according the “root” peaks of the corresponding ^{15}N-HSQC spectra. The so obtained strips are then matched in the following way:

For each assigned “old” strip (TM12) a number of potential “new” strips of the unassigned spectrum (TM127) are selected. This selection is based on the ^{1}H(F3) and ^{15}N(F2) shifts of the TM12 strip, defining a spectral window, within all potential new strips are searched. This window is larger for residues at the beginning and the tail of the sequence. For a residue at the terminal positions 1 or N the search window in the ^{1}H dimension is 0.25 ppm, and linearly decreases to 0.1 ppm over the next ten residues towards the center of the sequence. Similarly the search radius in the ^{15}N dimension decreases from 2.5 to 1 ppm. In the central segment the allowed shift deviation is 0.25 and 1 ppm for ^{1}H and ^{15}N for all residues.

The set of “new” strips are then compared to the “old” strip by calculating the differences of the “line-spectra”, and the “old” strip is matched to the “new” strip showing the smallest difference (see Figure 3.9). The differences are calculated by minimizing the following function:

$$f = \sum D_{ij} + \sum A_i + \sum B_j$$

where D_{ij} is the “distance” between line i of the “old” strip and line j of the “new” strip,

$$D_{ij} = c(\delta_i - \delta_j)^2 + (I_i - I_j)^2$$

with the chemical shift and the intensity of the line i and j, and c is a scaling factor. Note that the intensities have been normalized to the most intense line within a strip. Further we have

$$A_i = I_i^{0.3}$$

$$B_j = 0.5 \cdot I_j^2$$

which are the penalties for lines that are not included in the sum of D_{ij}’s.

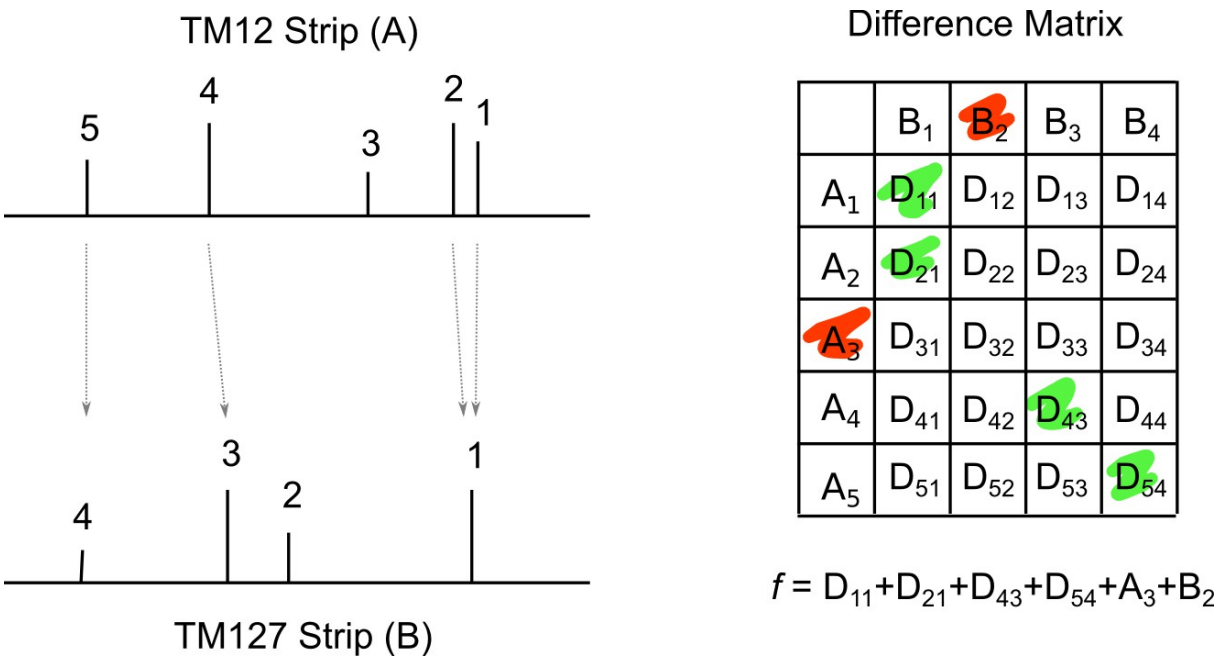


Figure 3.9: Schematic representation of the strip matching routine

Note that lines A3 and B2 are not matched, and hence A3 and B2 are added to the penalty function

The minimum of *f* is found in a systematic search routine, where in every row of the "difference matrix" (except the row with just the *B*'s) exactly one element is selected and calculating the sum after adding the *B*'s for empty columns. Note that in this way a peak of the "old" strip (A) may be simultaneously matched to two (or more) peaks in the "new" strip (B) while the opposite is impossible. It is also worth noticing that in practice only a small number of permutation must be calculated as the difference matrix can be much reduced by omitting all *D*'s for which $Dij>Ai+Bj$.

The constants in *f* are chosen emprically based on the success in matching the strips, and it is expected that can be further optimized.

3.4.8 Structure calculation

Distance restraints were obtained from ^{15}N-resolved NOESY spectra recorded on ^{15}N,^{1}H- and ^{15}N,^{2}H-labeled samples with mixing times of 70 and 120 ms, respectively, and from 100 ms ^{13}C-resolved NOESY spectra. In general, ^{15}N- or ^{13}C-resolved NOESY spectra were recorded with non-deuterated and deuterated detergents, respectively. In addition, dihedral angle restraints obtained using the program TALOS+[49], that uses chemical shifts of ^{1}Hα, ^{13}Cα, ^{13}Cβ, ^{13}C', and ^{15}N nuclei,

were added. Peak lists for each spectrum were picked by the program package UNIO'10[50, 51]. The automatically picked peaklists were manually edited to remove artifactual peaks (e.g. t_1 noise) or to pick additional weak peaks. The integrated, non-assigned peak list from UNIO'10 was subsequently transferred to CYANA[52], which annotated the peak list in seven iterative cycles using the build-in macro "noeassign". Again, the results from the automatic assignments were carefully checked and edited when necessary.

Peak positions in the ^{13}C- and ^{15}N-resolved NOESY spectra differed for samples recorded in deuterated from those in non-deuterated detergent. To account for that deviation chemical shift positions were related to the corresponding NOESY spectrum both manually and automatically using the sidechain adaptation routine of UNIO'10.

Table 3.1: NOE distance restraints

total	728	100.0%
Intraresidue |i-j |= 0	329	45.2%
sequential |i-j| = 1	283	39.0%
short-range |i-j| <= 1	612	84.2%
medium-range 1 < |i-j| < 5	103	14.2%
long-range |i-j| >= 5	13	1.7%

Due to the limited amount of long-range information (see Table 3.1) the global fold of TM127 could not be calculated based on NOE distance restraints and 166 torsion angle restraints alone. Therefore, a second round of structure calculations was performed using auxiliary sources of structural information available in the form of distance restraints derived from PRE measurements and restraints for residual dipolar couplings.

PRE distances were available in three classes as described above. It was therefore possible to include 26 PRE distances with an upper distance limit of 14Å and 46 PRE distances with a lower limit of 20Å. Additionally, 37 PRE distance restrained between 14Å and 20Å were defined by scaling them using an exponential component as introduced by Battiste *et al*[39] and implemented with bounds of ±4Å. PRE distance restraints required cross-checking for compliance with the short- to medium-range NOE restraint network, as well as the restraints for torsion angle restraints.

63 RDC restraints complemented the PRE-derived distance restraints. For the first iteration of structure calculations, artificial upper and lower distance limits were generated to incorporate structural information from soluble paramagnetic spin labels. The restraints were defined between the amide proton of residues with an attenuated signal due to the influence of the spin labels. The approach relied on assumptions made for the detergent micelle environment (detergent LPPG) based on reference literature[53]. The micelle was assumed to have an elongated form with a height of 38-40 Å and a length of 56.8-58.8 Å. Upon insertion of TM127 into the micelle in an orthogonal orientation, the center of the micelle was expected at 19-20 Å distance from the solvent-exposed loop regions. In order to represent the micelle nexus, a pseudo residue from the CYANA residue library was utilized. This residue was attached to the amino acid sequence of TM127 via flexible linker of sufficient length. The pseudo residue is insubstantial and cannot cause atom collisions.

Signals strongly influenced by the soluble spin labels and displaying a large degree (>80%) of attenuation received a lower limit of 19 Å to the micelle center. Signals that were not subjected to line broadening resulting from the spin label influence were assumed to be more deeply embedded within the micelle and restrained with upper an upper limit of 15 Å. Signal attenuation was also observed using HNCO experiments. In order to apply information from these measurements to the structure calculation, ambiguous distance restraints were created for atoms H and N of the residue corresponding to an observed attenuated signal, as well as for the carbonyl C of the predecessor amino acid with the aforementioned upper and lower distance limits of 15 Å and 19 Å, respectively. However, the incorporation of distance restraints derived from soluble PREs led to over-restraint structures. The uneven number of observable PREs from soluble spinlabels for different regions in the protein caused the artificially created micelle center to be located outside of the perceived helix bundle. This led to very strong constraints, even at reduced weights, on certain parts of the construct, effectively 'bending' the helices. Restraints from soluble spin labels were therefore omitted from further structure calculations and rather used for qualitative validation for the structures. In total, the restraints shown in Table 3.2 were applied in the structure calculation of TM127

Table 3.2: Structure calculation statistics for TM127

Distance restraints		
NOE distance restraints		726
Torsion angle restraints		166
PRE-derived distance restraints	Upper limit only	26
	Upper and lower limit	37
	Lower limit only	47
RDC restraints		63
Structure calculation statistics		
Ave. target function		7.19 $Å^2$
Ave. backbone RMSD to mean		4.53 Å
Structural clusters		48

In the final structure calculation, 5000 structures were calculated in 10 individual runs with 500 structures each from 10 different random seeds using standard structure calculation routine in CYANA with 20000 torsion angle dynamics steps. The 50 structures with the lowest target function were collected from each run and combined to a bundle containing 500 structures, which was clustered according to the backbone RMSD of helix regions T50-T72, P79-L102 and V276-A299 with a cutoff of 3Å. Clustering resulted in 48 total clusters with the largest containing 95 structures and 11 clusters containing at least 10 structures. 16 clusters with only 1 structure, 5 clusters with 2 and 4 clusters with three structures could be observed.

3.5 Discussion

Despite the pivotal role of GPCRs in biology and despite its importance in pharmacology and pharmaceutical sciences structural data of GPCRs in solution are still lacking. Here were have presented an approach based on the study of a series of overlapping fragments of extended size that aims at providing methodology of assigning entire receptors, that may possibly be more generic for the study of α-helical membrane proteins. Moreover, we suggest that the system provides valuable insight into conformational preferences of protein segments as they are formed during early stages of protein biosynthesis.

During structural studies of membrane proteins by solution NMR obtaining chemical shift assignments presents a significant general bottleneck. Here we demonstrated that the knowledge of chemical shifts from fragments can be utilized to assign larger proteins for which such assignments are difficult to obtain otherwise. We suggest a method for chemical shift adaptations that requires $^{15}N,^{1}H$ correlation maps for the fragment and the entire receptor as well as the corresponding 3D ^{15}N-NOESY spectra. Surprisingly, we could correctly transfer more than 90% of the assignments from TM12 to TM127, and during the adaptation process also sidechain assignments are transferred. To which extent this method will be also successful for transferring assignments onto the entire receptor will depend on a number of factors, most important being the quality of the ^{15}N-NOESY spectrum of the entire receptor (essentially how much NOEs are contained in the strips), and by which extent sidechain chemical shifts change upon formation of tertiary structure. The quality of the NOESY spectrum depends on the thermal stability of the protein in the chosen membrane-mimicking environment, and we expect that temperature-stabilized forms may still display acceptable spectra. In cases where large portions of signals in the $[^{15}N,^{1}H]$-TROSY spectrum are missing (e.g. those from the TM segments) success perspectives for *any* NMR experiments that go beyond 1D spectroscopy will become small. We like to note here that the ^{15}N-NOESY experiment is usually the one that works the best even in case of large proteins. Some of the sidechain resonances will of course change upon formation of tertiary structure. The number of residues, for which *new* sidechain contacts are formed, will be limited though. In any case they usually do not give rise to new NOEs since the distance to amide protons of adjacent helices is too large, but of course such contacts result in chemical shift changes of the involved sidechain proteins. There is still a reasonable chance that the chemical shift pattern does not change too much, so that it can be matched in combination with knowledge of to which type of amino acid the amide moiety belongs, obtained from amino-acid selectively labeled samples.

The main focus of this study was to investigate whether addition of one or more helices to TM1 that in the entire receptor form contacts with it help to stabilize it. Proteins are synthesized starting at the N-terminus, and the co-translational folding model postulates that folding occurs after the nascent chain is released via the translocon into the membrane compartment[13, 54]. The two-stage model for folding helical membrane proteins proposes that secondary structure forms once the entire TM segment partitions into the membrane[11, 12]. Helical TM-bundle proteins, however, contain polar or even charged residues at TM-internal positions. These form contacts with polar or charged groups in other TM helices, and thereby contribute to the assembly of the TM-bundle, and in particular to the specificity of interhelical contacts[21, 55, 56]. Insertion of single TM helices into the hydrophobic core exposes these polar moieties to the lipids, an unfavorable interaction that is expected to expel the corresponding part from the membrane interior and transfer it into the interface. Indeed, we observed that the first helix is significantly destabilized in TM1, and in fact interrupted around the central polar GVRSG motif. Packing of TM2 against TM1 apparently stabilizes secondary structure in TM1 as demonstrated by the fact that TM12 forms a much more stable α-helical hairpin. We expected that adding another TM helix would even stabilize the fragment further. In fact, we observed formation of a comparably stable hydrophobic core in TM127 (but not in TM123). Our NMR data, however, indicate that TM127 does not exist in one well-defined conformation. In particular, hinge motions about Pro-290 in TM7 are seen. The nearly parallel orientation of the TM helices predicted from homology modeling based on the rhodopsin structure[57] is clearly incompatible with the NMR data, and TM2 packs almost orthogonal against TM1 and TM7. The presence of exchange-broadened amide resonances as well as the fact that the interhelical PRE effects are smaller than predicted indicate that either the interhelical packing is not stable, or that (slightly) differently packed conformers interchange. Interestingly, we can exclude any form in which the individual helices present lipid-associated but well-separated surface-associated entities despite the presence of various polar or even charged residues at locations central in the TM helix. The Gd-DPTA-BMA data indicated that solvent access is mostly limited to the loops, indicating that the TM helices of TM127 are integrated into the hydrophobic core. In contrast, TM3 in TM123 displays significant solvent access in the center of TM3. To conclude, TM127 in detergent micelles presents a loosely packed bundle of 3 TM helices, all of which are well-integrated into the hydrophobic core, with non-native topology. Although we cannot present a structure of TM123 at the moment, it seems structurally to be more inhomogeneous, exists in various oligomeric states, and TM3 is not completely integrated into the hydrophobic core.

Recently we have performed an exhaustive study on folding of N- and C-terminally truncated forms of the Y4 receptor, a human GPCR[24]. In these studies the green-fluorescent protein (GFP) or the alkaline phosphatase (phoA) were added as reporters to the C-terminus, and proteins expressed in the *E.coli* inner membrane using a biochemical readout developed by the van Heijne lab[58]. The data from this study indicated that dual topologies (C-terminus in *and* out) are more likely occur for the short fragments (e.g. TM1 or TM12) while more unique and correct topologies were encountered for proteins that comprise most of the TM helices (e.g. TM16). Again, we suspected that the presence of uncompensated polar or charged residues at central position in the TM helices is responsible for that behavior.

What happens to helices that do not readily insert into the membrane because polar or charged residues prevent insertion? In case of a sequential exit from the translocon such helices will localize in the interface, at least to some extent. For example, our Gd-DPTA-BMA data indicate that central parts of TM3 have solvent access, and hence TM3 is not fully inserted into the membrane. Likely, its full membrane insertion requires that compensating interactions from other, later synthesized, helices occur. Alternatively, nascent helices may accumulate within or near the translocon, were the bundle is already assembled and then fully released into membrane. Photo-crosslinking experiments demonstrated that TM helices formed either specific or non-specific contacts with residues from the Sec61 translocon[59]. Interestingly, the experiments also revealed that the TM helices form contacts with both the translocon as well as with the phospholipids. Ismail and coworkers using similar methodology demonstrated that large portions of opsin remain bound or associated to the translocon[60]. The rate by which they were released depended more on the amino acid sequence of the TM helix rather than by its position in the protein. In their model the nascent chain remains within the translocon or attached to translocon-associated other proteins until the protein has been fully synthesized (for a more general discussion of the issue see also the review by Skach[13]).

In the experiments described in this work we have probed for conformational preferences of N-terminal fragments of extended length of a yeast GPCR. Thereby we were able to obtain experimental data about how these protein fragments would behave if there were released into the hydrophobic core from the translocon. The data basically support conclusions from Ismail *et al.*[60] in that the rate of release is dictated by the amino acid sequence of the TM, and in particular by the hydrophobicity of the segment. It thus seems more likely that larger portions of the receptor remain either within the translocon or associated with nearby proteins until tertiary contacts between polar or charged residues are made. Otherwise, some of these TM helices simply will not remain

integrated into the membrane but rather accumulate in the interface. The exact folding pathway therefore seems to depend very much on the amino acid sequence, and hence may vary from receptor to receptor.

ACKNOWLEDGMENT

We thank the Swiss National Science Foundation (grant no. 31003A_124469) for funding. Fred Naider acknowledges support from NIGMS Grant 22087 and the Leonard and Esther Kurtz Term Professorship. Thank to Gunnar Jeschke for EPR measurements.

3.6 References

1. Neumoin A, Cohen L, Arshava B, Tantry S, Becker J, Zerbe O, et al. *Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles.* Biophys J. 2009;**96:**3187-96.
2. Lefkowitz RJ. *A brief history of G-protein coupled receptors (Nobel Lecture).* Angew Chem Int Ed Engl. 2013;**52:**6366-78.
3. Kobilka B. *The structural basis of G-protein-coupled receptor signaling (Nobel Lecture).* Angew Chem Int Ed Engl. 2013;**52:**6380-8.
4. Venkatakrishnan A, Deupi X, Lebon G, Tate C, Schertler G, Babu M. *Molecular signatures of G-protein-coupled receptors.* Nature. 2013;**494:**185-94.
5. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science. 2000;**289:**739-45.
6. Rasmussen S, Choi H, Rosenbaum D, Kobilka T, Thian F, Edwards P, et al. *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature. 2007;**450:**383-7.
7. Ahlquist RP. *A study of the adrenotropic receptors.* The American journal of physiology. 1948;**153:**586-600.
8. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, et al. *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science. 2007;**318:**1258-65.
9. Rasmussen S, DeVree B, Zou Y, Kruse A, Chung K, Kobilka T, et al. *Crystal structure of the beta2 adrenergic receptor-Gs protein complex.* Nature. 2011;**477:**549-55.
10. Popot JL, Berry EA, Charvolin D, Creuzenet C, Ebel C, Engelman DM, et al. *Amphipols: polymeric surfactants for membrane biology research.* Cell Mol Life Sci. 2003;**60:**1559-74.
11. Popot JL, Engelman DM. *Membrane protein folding and oligomerization: the two-stage model.* Biochemistry. 1990;**29:**4031-7.
12. Popot JL, Engelman DM. *Helical membrane protein folding, stability, and evolution.* Annu Rev Biochem. 2000;**69:**881-922.
13. Skach W. *Cellular mechanisms of membrane protein folding.* Nat Struct Mol Biol. 2009;**16:**606-12.
14. Wimley WC, White SH. *Experimentally determined hydrophobicity scale for proteins at membrane interfaces.* Nature Struct Biol. 1996;**3:**842-8.
15. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, et al. *Molecular code for transmembrane-helix recognition by the Sec61 translocon.* Nature. 2007;**450:**1026-30.
16. White S, von Heijne G. *How translocons select transmembrane helices.* Annu Rev Biophys. 2008;**37:**23-42.
17. White SH. *Translocons, thermodynamics, and the folding of membrane proteins.* FEBS Lett. 2003;**555:**116-21.
18. Bowie JU. *Structural biology. Membrane protein twists and turns.* Science. 2013;**339:**398-

9.

19. Virkki MT, Agrawal N, Edsbacker E, Cristobal S, Elofsson A, Kauko A. *Folding of Aquaporin 1: multiple evidence that helix 3 can shift out of the membrane core.* Protein Sci. 2014;**23:**981-92.

20. Lu H, Booth PJ. *The final stages of folding of the membrane protein bacteriorhodopsin occur by kinetically indistinguishable parallel folding paths that are mediated by pH.* J Mol Biol. 2000;**299:**233-43.

21. Senes A, Engel DE, DeGrado WF. *Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs.* Curr Opin Struct Biol. 2004;**14:**465-79.

22. Gratkowski H, Lear JD, DeGrado WF. *Polar side chains drive the association of model transmembrane peptides.* Proc Natl Acad Sci U S A. 2001;**98:**880-5.

23. Chamberlain AK, Faham S, Yohannan S, Bowie JU. *Construction of helix-bundle membrane proteins.* Adv Protein Chem. 2003;**63:**19-46.

24. Marino J, Geertsma ER, Zerbe O. *Topogenesis of heterologously expressed fragments of the human Y4 GPCR.* Biochimica et Biophysica Acta (BBA) - Biomembranes. 2012;**1818:**3055-63.

25. Shao X, Zou C, Naider F, Zerbe O. *Comparison of fragments comprising the first two helices of the human Y4 and the yeast Ste2p G-protein-coupled receptors.* Biophys J. 2012;**103:**817-26.

26. Staley JP, Kim PS. *Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor.* Protein Science : A Publication of the Protein Society. 1994;**3:**1822-32.

27. Grisshammer R, Duckworth R, Henderson R. *Expression of a rat neurotensin receptor in Escherichia coli.* Biochemical Journal. 1993;**295:**571-6.

28. Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. *Conformational thermostabilization of the 1-adrenergic receptor in a detergent-resistant form.* Proc Natl Acad Sci U S A. 2008**:** .

29. Chowdhury A, Feng R, Tong Q, Zhang Y, Xie X-Q. *Mistic and TarCF as fusion protein partners for functional expression of the cannabinoid receptor 2 in Escherichia coli.* Protein expression and purification. 2012;**83:**128,Äì34.

30. Roosild T, Greenwald J, Vega M, Castronovo S, Riek R, Choe S. *NMR structure of Mistic, a membrane-integrating protein for membrane protein expression.* Science (New York, NY). 2005;**307:**1317,Äì21.

31. Cohen LS, Becker JM, Naider F. *Biosynthesis of peptide fragments of eukaryotic GPCRs in Escherichia coli by directing expression into inclusion bodies.* Journal of Peptide Science. 2010;**16:**213-8.

32. Martin N, Celic A, Dumont M. *Mutagenic mapping of helical structures in the transmembrane segments of the yeast alpha-factor receptor.* J Mol Biol. 2002;**317:**765-88.

33. Caroccia K, Estephan R, Cohen L, Arshava B, Hauser M, Zerbe O, et al. *Expression and biophysical analysis of a triple-transmembrane domain-containing fragment from a yeast G protein-coupled receptor.* Biopolymers. 2011;**1808:**2674-84.

34. Gottstein D, Reckel S, Dötsch V, Güntert P. *Requirements on Paramagnetic Relaxation Enhancement Data for Membrane Protein Structure Determination by NMR.* Structure. 2012;**20:**1019-27.

35. Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. *A suite of triple-resonance NMR experiments for the backbone assignment of N-15,C-13,H2- labeled proteins with high sensitivity* J Am Chem Soc. 1994;**116:**11655-66.

36. Shan X, Gardner KH, Muhandiram DR, Rao NS, Arrowsmith CH, Kay LE. *Assignment of N-15, C-13(alpha), C-13(beta), and HN resonances in an N-15, C-13, H-2 labeled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and H-2-decoupling.* J Am Chem Soc. 1996;**118:**6570-9.

37. Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD. *A gradient-enhanced HCCH-TOCSY experiment for recording side-chan 1H and 13C correlations in H2O samples of proteins.* J Magn Reson Ser B. 1993;**101:**333-7.

38. Olejniczak ET, Xu RX, Fesik SW. *A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins.* J Biomol NMR. 1992;**2:**655-9.

39. Battiste JL, Wagner G. *Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data†.* Biochemistry. 2000;**39:**5355-65.

40. Geertsma E. *FX cloning: a simple and robust high-throughput cloning method for protein expression.* Methods in molecular biology (Clifton, NJ). 2014;**1116:**153,Äì64.

41. Tugarinov V, Kay LE. *Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods.* J Am Chem Soc. 2003;**125:**13868-78.

42. Cohen LS, Arshava B, Estephan R, Englander J, Kim H, Hauser M, et al. *Expression and biophysical analysis of two double-transmembrane domain-containing fragments from a yeast G protein-coupled receptor.* Biopolymers. 2008;**90:**117-30.

43. Killian JA, Trouard TP, Greathouse DV, Chupin V, Lindblom G. *A general method for the preparation of mixed micelles of hydrophobic peptides and sodium dodecyl sulphate.* FEBS Lett. 1994;**348:**161-5.

44. Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE. *Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of 13C labelled proteins.* J Magn Reson. 1990;**87:**620-7.

45. Bax A, Clore M, Gronenborn AM. *1H-1H Correlation via Isotropic Mixing of 13C Magnetization, a New Three-Dimensional Approach for Assigning 1H and 13C Spectra of 13C Enriched Proteins.* J Magn Reson. 1990;**88:**425-31.

46. Diercks T, Coles M, Kessler H. *An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments.* J Biomol NMR. 1999;**15:**177-80.

47. Keller R. *The Computer Aided Resonance Assignment.* 1252.

48. Dosset P, Hus JC, Marion D, Blackledge M. *A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings.* J Biomol NMR. 2001;**20:**223-31.

49. Shen Y, Delaglio F, Cornilescu G, Bax A. *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR. 2009;**44:**213-23.

50. Herrmann T, Güntert P, Wüthrich K. *Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.* J Mol Biol. 2002;**319:**209-27.

51. Herrmann T, Güntert P, Wüthrich K. *Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS.* J Biomol NMR. 2002;**24:**171-89.

52. Güntert P. *Automated NMR structure calculation with CYANA.* Methods Mol Biol. 2004;**278:**353-78.

53. Lipfert J, Columbus L, Chu VB, Lesley SA, Doniach S. *Size and shape of detergent micelles determined by small-angle X-ray scattering.* J Phys Chem B. 2007;**111:**12427-38.

54. Rapoport TA. *Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes.* Nature. 2007;**450:**663-9.

55. Zhou FX, Merianos HJ, Brunger AT, Engelman DM. *Polar residues drive association of polyleucine transmembrane helices.* Proc Natl Acad Sci USA. 2001;**98:**2250-5.

56. Choma C, Gratkowski H, Lear JD, DeGrado WF. *Asparagine-mediated self-association of a model transmembrane helix.* Nat Struct Biol. 2000;**7:**161-6.

57. Eilers M, Hornak V, Smith S, Konopka J. *Comparison of class A and D G protein-coupled receptors: common features in structure and activation.* Biochemistry. 2005;**44:**8959-75.

58. Drew D, Sjostrand D, Nilsson J, Urbig T, Chin Cn, Gier JWd, et al. *Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis.* Proceedings of the National Academy of Sciences of the United States of America. 2002;**99:**2690-5.

59. McCormick PJ, Miao Y, Shao Y, Lin J, Johnson AE. *Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins.* Molecular cell. 2003;**12:**329-41.

60. Ismail N, Crawshaw SG, Cross BC, Haagsma AC, High S. *Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis.* Biochem J. 2008;**411:**495-506.

3.7 Supplemental Material

3.7.1 Table of Contents

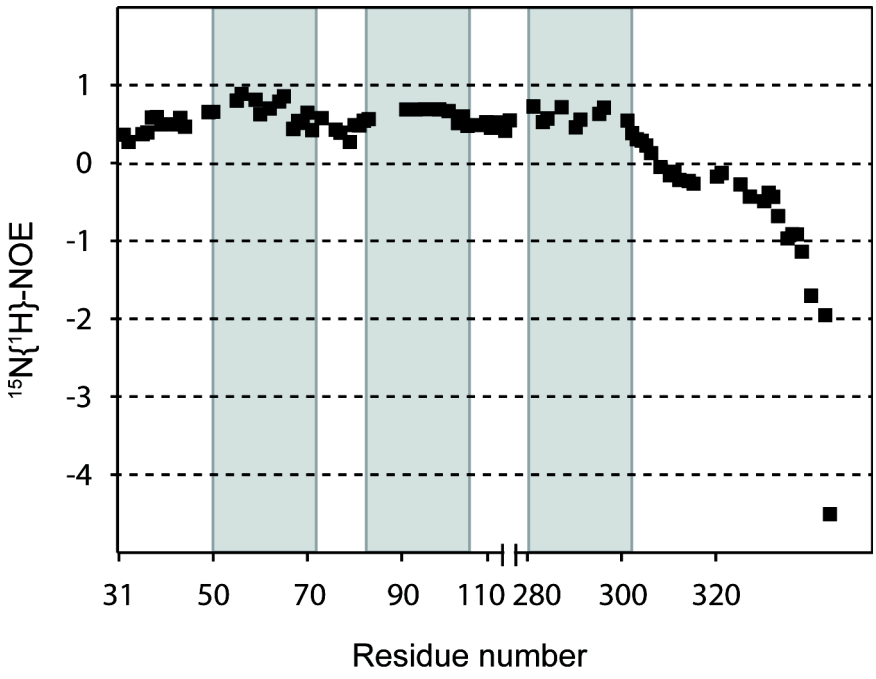


Figure S3.10: Het-NOE of TM127

$^{15}N\{^{1}H\}$-NOE of TM127, measured at 317K. The N-terminus appears to interact with the micelle, whereas the C-terminus seems completely unstructured.

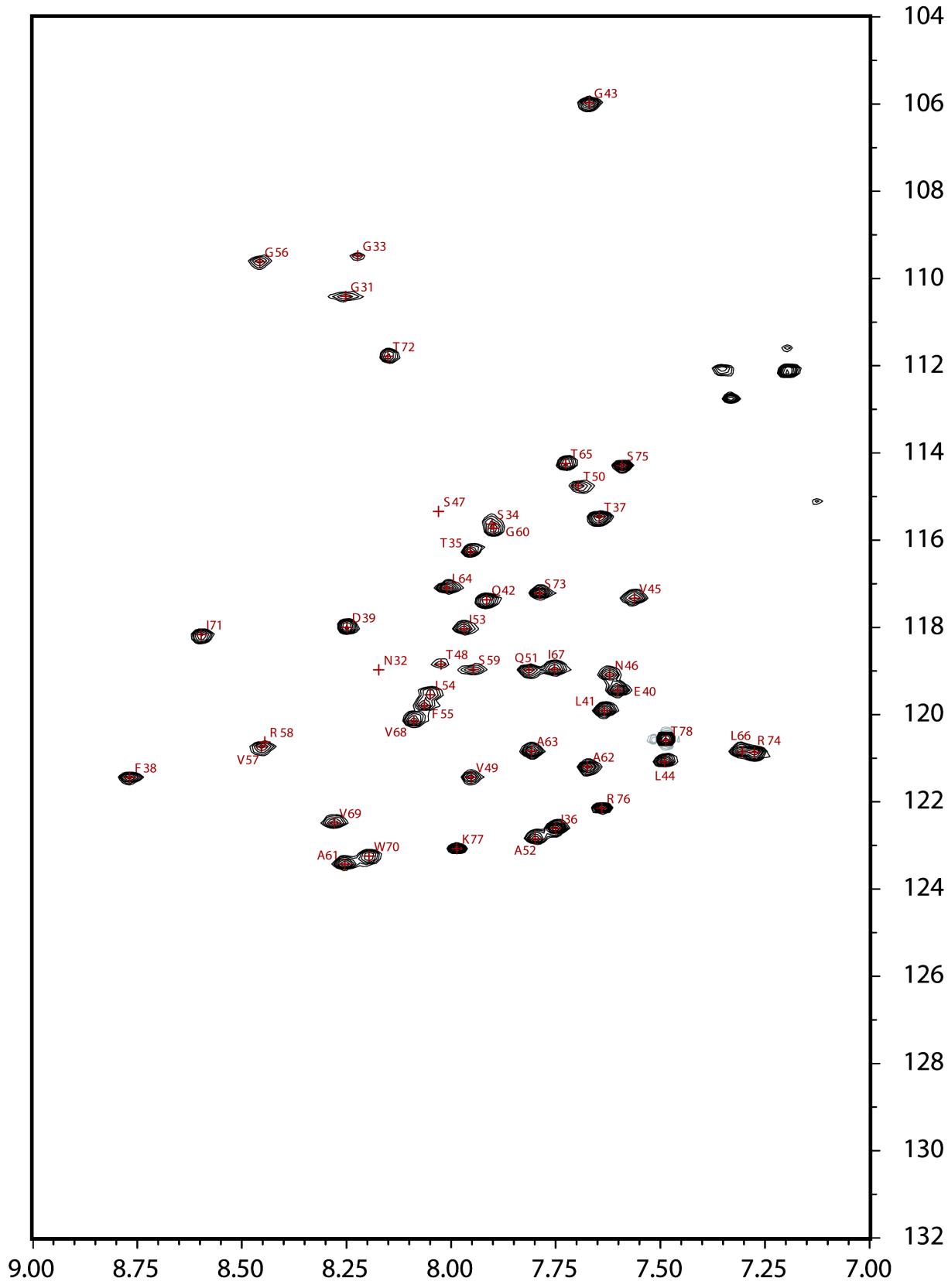


Figure S3.11:[^{15}N,^{1}H]-HSQC of TM1

TM1 exhibits homogeneous linewidths. 97% of the backbone could be assigned.

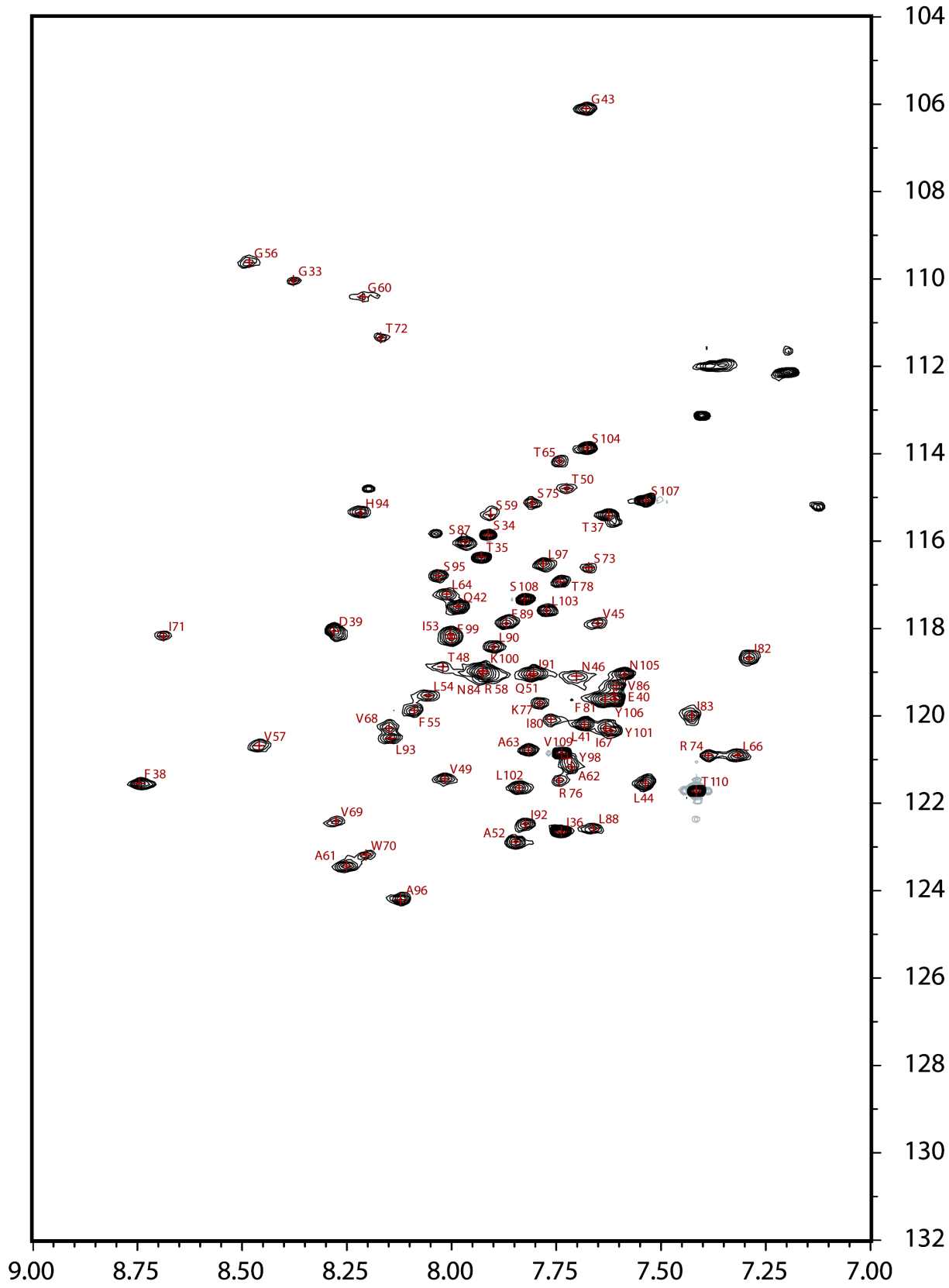


Figure S3.12: [^{15}N,^{1}H]-HSQC of TM12

TM12 exhibits homogeneous linewidths. 95% of the backbone could be assigned.

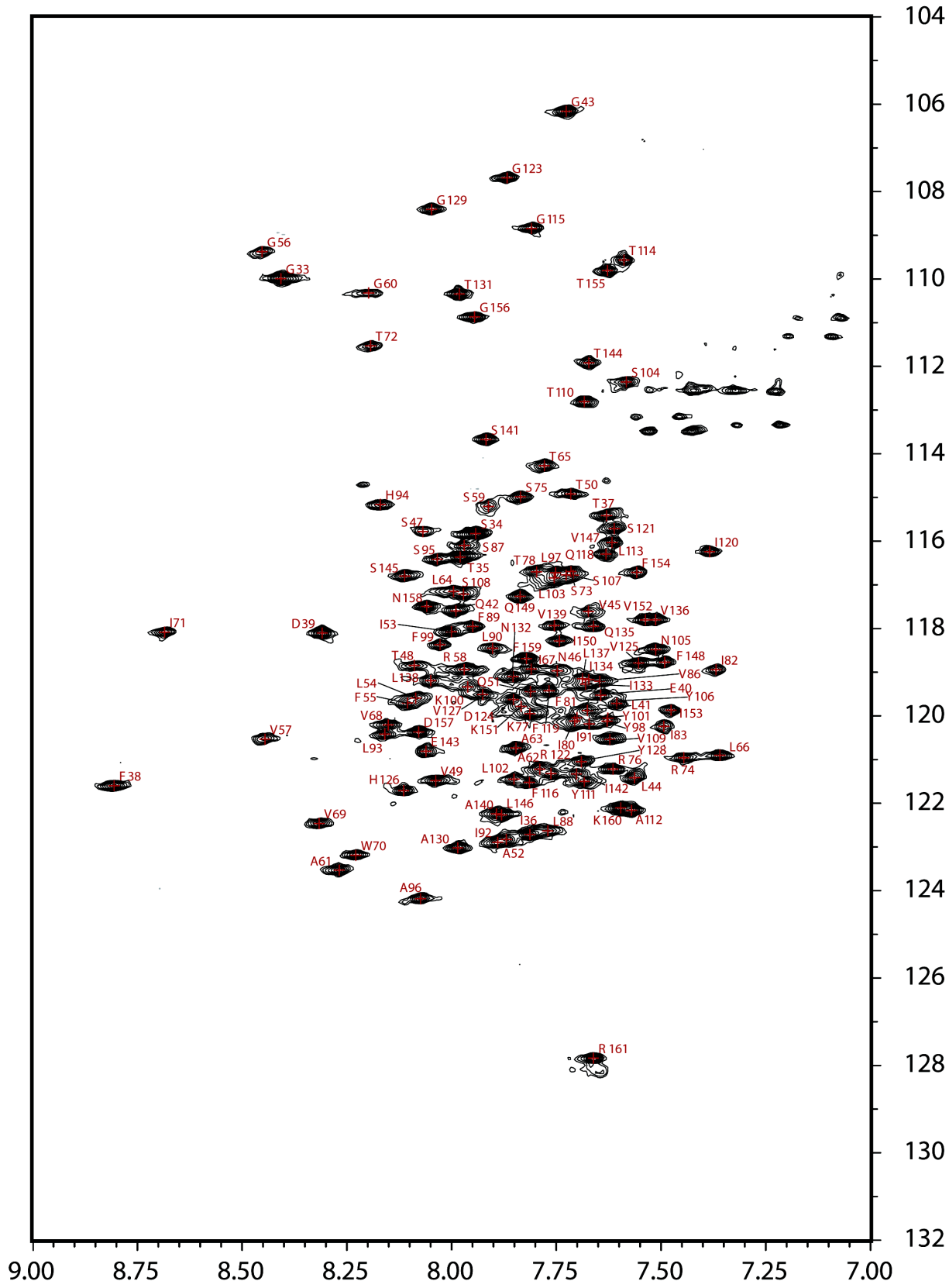


Figure S3.13: [^{15}N,^{1}H]-HSQC of TM123

TM123 exhibits some line-broadening in the center region. 95% of the backbone could be assigned.

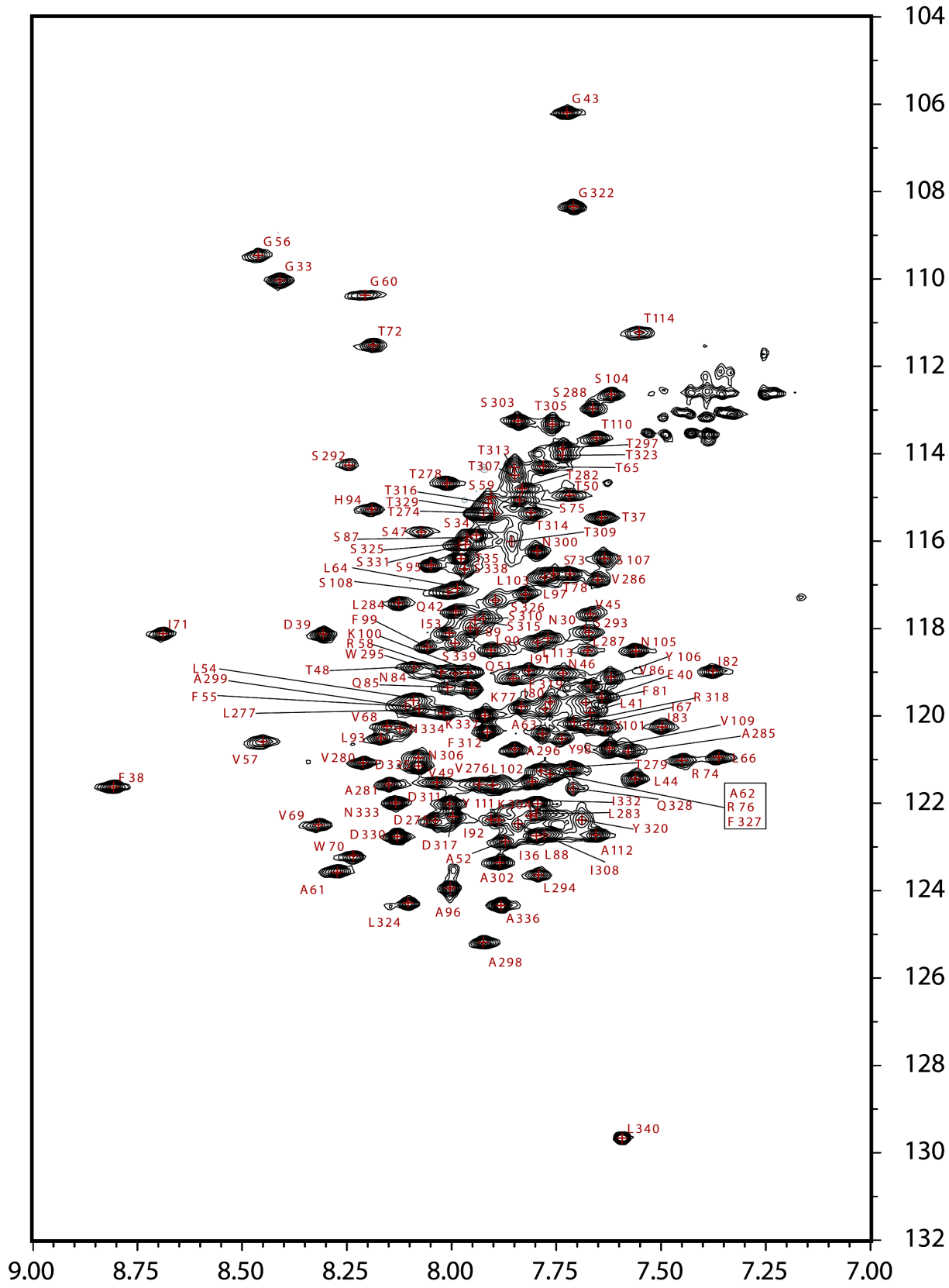


Figure S3.14: [^{15}N,^{1}H]-HSQC of TM127

TM127 exhibits homogeneous line-shapes. 93% of the backbone could be assigned

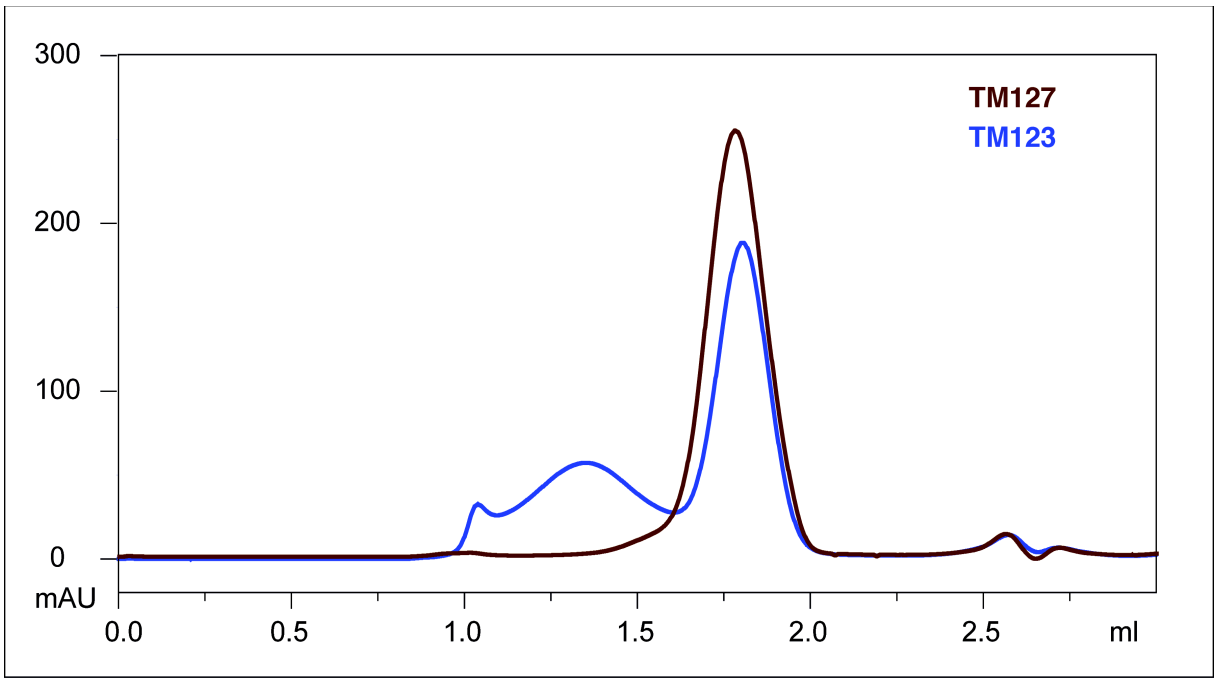


Figure S3.15: SEC overlay of TM123 and TM127

Freshly prepared TM123 (blue) already forms large soluble aggregates, whereas TM127 (brown) is mostly monomeric. A significant amount of TM123 can already be found in the void volume at around 1.0ml and in the broad peak from 1.1ml to 1.6ml. SEC runs were performed on an analytical S75 column with a flowrate of 0.5ml/min 40mM phosphate buffer, containing 15mM LPPG/DPC (4:1) and 150mM NaCl at pH 6.5. The sample buffer contained 150mM LPPG/DPC (4:1) with a protein concentration of 200µM.

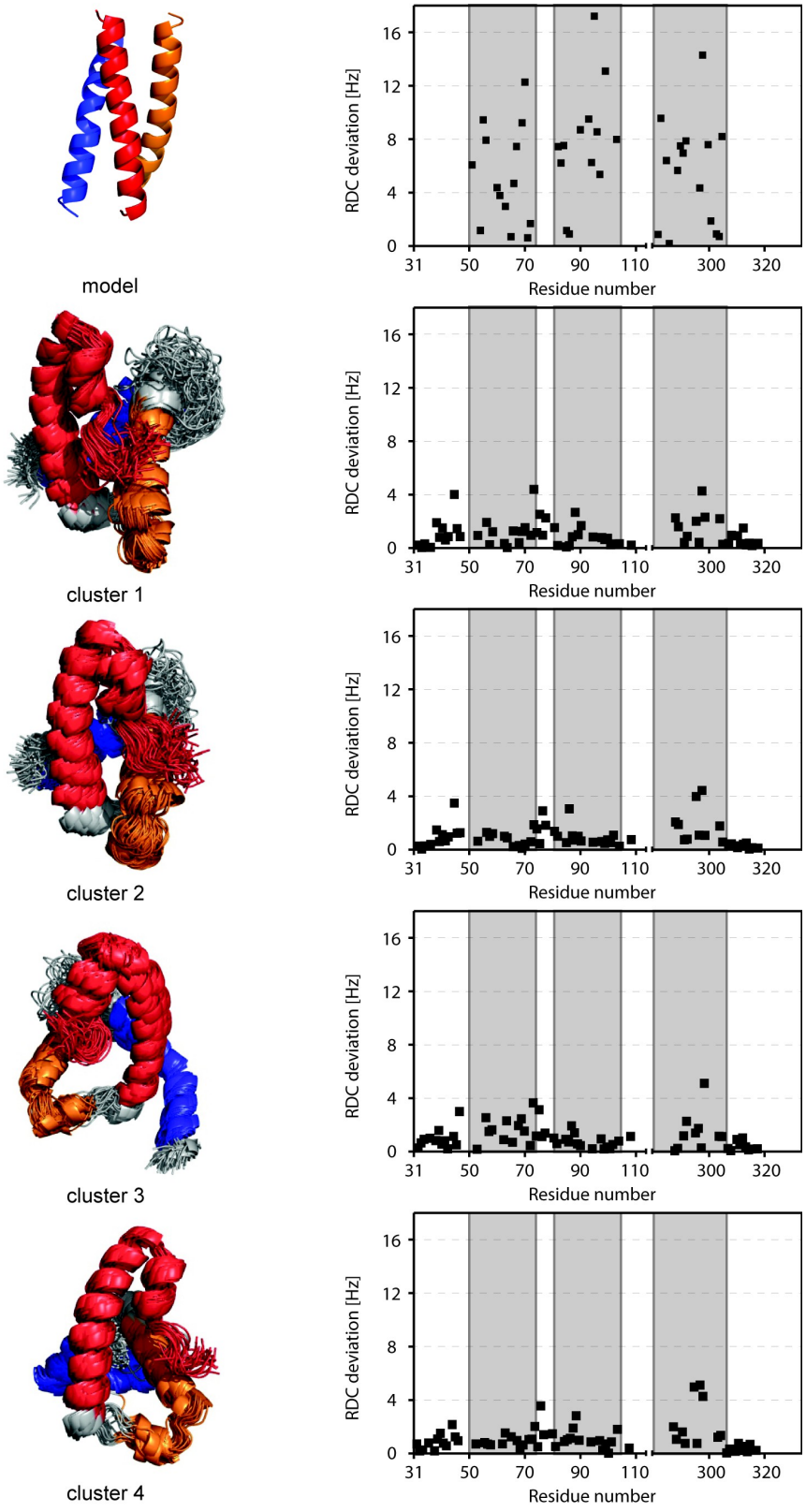


Figure S3.16: Comparison of simulated RDCs

Difference of simulated versus experimental RDCs for the four most prominent clusters as well as the homology model. RDCs based on the model are clearly incompatible with the experimental data.

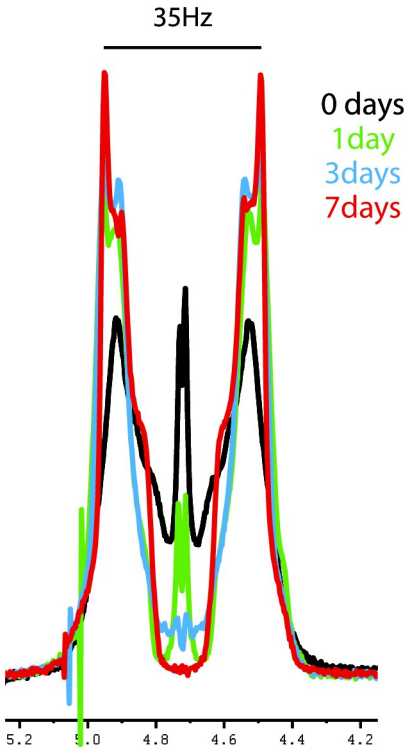


Figure S3.17: Maturation of the alignment phase as seen in the deuterium splitting.

Deuterium splitting of dGpG RDC phases. Several heating and cooling cycles (4° to 45°) were necessary over the course of one week to completely dissolve the dinucleotides and achieve acceptable deuterium splitting. In presence of detergents, the dissolution of dGpG is severely impeded, but can be accelerated using high concentrations of nucleotides (in this case 30mg/ml), high potassium ion concentrations (100mM) and the implementation of heating-cooling cycles.

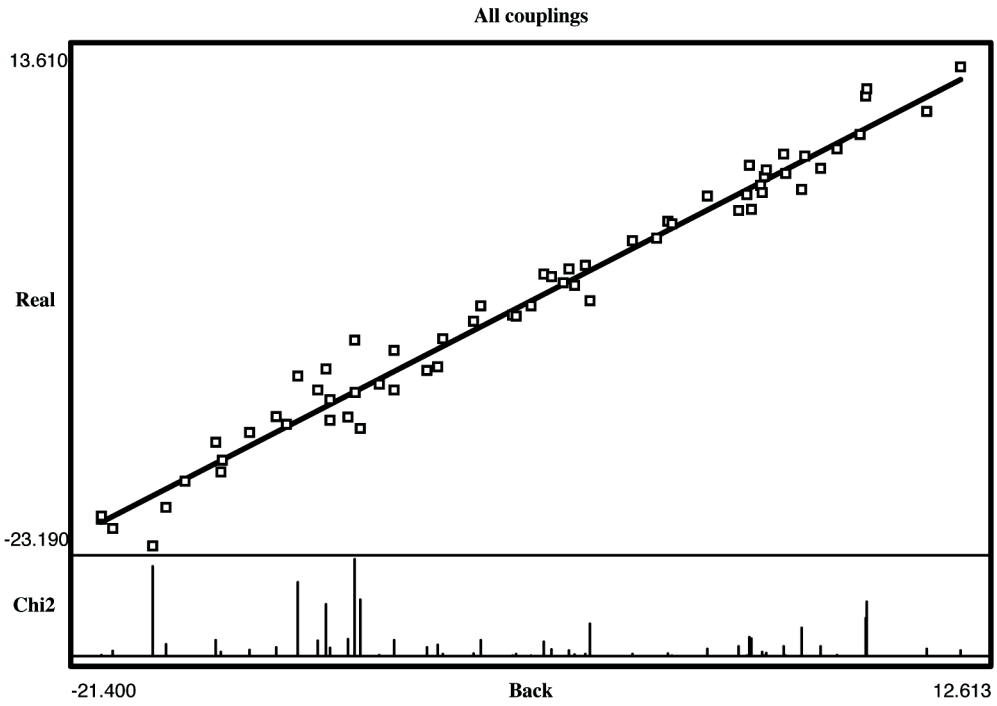


Figure S3.18: Validation of NH-RDC values using Single Value Decomposition (SVG).

RDCs were simulated using the program MODULE and compared to experimental data. The bottom part shows Chi2 values.

4 Understanding GPCR structures and folding from large fragments

J. Marino, M. Poms[1], O. Zerbe[1]*

[1]Department of Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057

*Corresponding author:
Oliver Zerbe
Department of Chemistry
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland
phone: +41-44-6354263
fax: +41-44-6356882
email: oliver.zerbe@chem.uzh.ch

4.1 Introduction

4.1.1 Understanding GPCR structure and folding from large fragments

G-protein coupled receptors (GPCRs) constitute a very important class of integral membrane proteins that help transmitting the signal from the outside to the inside of cells. Accordingly, these proteins pay a pivotal role in biology, and more than 30% of all commercial drugs are believed to act via binding to GPCRs.

The important role of GPCR has triggered lots of efforts to determine the structure of these precious receptors. First low-resolution data emerged from cryo-electron microscopy[1] but the scientific community had to wait for the first crystal structure, bovine rhodopsin, until 2000[2]. In this report rhodopsin was isolated from the retina, and it took 6 more years until the first GPCR produced from recombinant sources, the structure of the β_2-adrenergic receptor, was released[3]. Since then we have witnessed publication of new structures at an ever-accelerating pace culminating in the awarding of the Nobel Prize in chemistry to Brian Kobilka and Bob Lewkowitz in 2012[4].

Structural studies of GPCRs are largely hampered by two major issues[5]: the recombinant production of functional GPCRs at sufficient yields is very difficult, and many receptors have proven to be not amenable to structural studies. Most GPCRs that were structurally characterized have been produced in insect cells, but also mammalian, yeast or *E. coli* cells were used. The second issue is related to the flexibility of the receptors that is partly related to the way they are activated. In particular most receptors comprise a rather long and flexible third cytosolic loop.

Methods for producing detergent- and temperature-stabilized mutants in *E. coli* have been introduced[6]. Initially, such mutants we obtained from a systematic mutagenesis screen, first preparing single mutants and then combining multiple beneficial mutations[7]. Recently, a clever system was introduced, in which the cells were coated with a chitin shell, so that cell content was not released upon addition of detergents[8]. Fluorescent ligand in combination with FACS sorting allowed screening for mutants that still bind receptors in presence of detergents (detergent-resistent mutants), making the system amenable to work with large libraries. As a result expression yields, and, even more important, temperature stability could be improved. For the purpose of solution NMR studies it is of pivotal importance to improve the latter two parameters.

Despite all the progress seen in crystallography, no solution-NMR GPCR structure has been published yet. Nevertheless, the seminal paper by the Nietlispach group reporting on sensory rhodopsin[9] and the structure of proteorhodopsin[10] from the Frankfurt NMR groups, although both proteins strictly speaking do not represent true GPCRs, have indicated that solution NMR in principle is capable of determining the structures of these precious membrane proteins provided they are well-behaved. However, both proteo- and sensory-rhodopsin present fairly rigid 7-TM proteins when compared with true GPCRs. In the latter the mode of activation requires conformational changes, and the inherent flexibility for that task results in conformational broadening of resonances. Much of the line-broadening that is present of spectra of GPCRs is likely not due to efficient relaxation due to slow overall reorientation but rather due to conformational broadening. Successful NMR studies of GPCRs may therefore require working with temperature-stabilized mutants that are also selected for improved detergent stability[5].

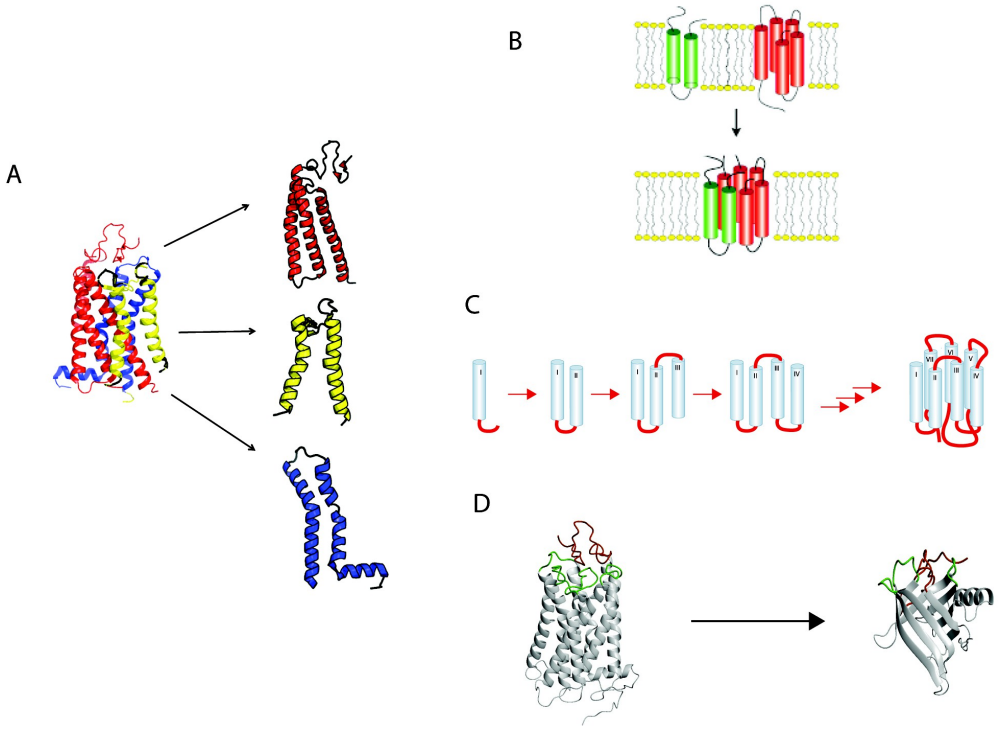


Figure 4.1: Approaches to study GPCRs from fragments pursued in our group.

A) The entire receptor is dissected into 2-3 TM fragments. B) Complementary fragments are differentially labeled and reconstituted into a split receptor. C) Fragments of increasing length are studied to learn about GPCR folding. D) Loops of a GPCR are grafted onto a β-barrel protein

Our group has started working on large fragments of GPCRs almost ten years ago. We decided to work on fragments because we thought they were easier to biosynthesize. Moreover, solution NMR, and in particular the process of sequence-specific resonance assignments, was expected to be facilitated.

Our strategy for using fragments is depicted in Figure 4.1. First of all, we like to study fragments to learn about conformational preferences of these polypeptide stretches (Figure 4.1A). Using complementary fragments reconstituting a split-receptor would allow us to determine the conformational preferences of these fragments (Figure 4.1B), however, this time in the context of a functional receptor, also providing a convenient way to make segmentally labeled receptors. As described in more detail below, we arealso convinced that we can learn about the folding of GPCRs from a study of conformational preferences of systematically truncated fragments (Figure 4.1C). We are additionally interested to know whether we can transfer chemical shift information from these fragments to the entire receptors to eventually enable determination of the structure of the entire receptor. Finally we have also tried to transfer the loops onto a well-behaved scaffold in order to create a mimic that can bind the ligands (Figure 4.1D).

4.1.2 General remarks on studying protein fragments

Does it make sense to look at protein fragments? This question is much disputed. For soluble proteins this is usually not the case. Dissecting a protein within domain borders will usually result in unstructured and insoluble aggregates, because not all necessary tertiary contacts can be made and hydrophobic residues are exposed to solvent. For membrane proteins this is a little different. Secondary structure formation is promoted by the hydrophobic environment, because only through formation of hydrogen bonds the polar peptide bond is shielded from making unfavorable contacts to lipids[11, 12].

The popular two-stage membrane protein folding model proposed by Popot and Engelman has postulated that segments of membrane proteins form secondary structure upon partitioning into the water-membrane interface[13]. These helices subsequently insert into the hydrophobic core of the membrane and diffuse within the bilayer until contacts with other helices are made eventually leading to the correct assembly of the helical bundle. If secondary structure in membrane environments is predominantly determined by sequence and does not depend on the formation of tertiary contacts, then it should be retained in fragments. As described below we usually observe

good agreement of the location of helical segments with predictions based on homology models, indicating that secondary structure of membrane proteins is primarily encoded in the amino acid sequence.

How about tertiary structure? A number of features have proposed to be important for the formation of interhelical contacts[14-17]. Those are for example shape-complementary amino acids, e.g. the knobs-in-hole arrangement in GXXXG motifs[18, 19]. Another way to mediate helical contacts is through formation of aromatic sidechains interactions[20] or via formation of hydrogen bonds[21]. The latter requires the presence of polar residues in sequence locations where they would not be expected based on a match of hydrophobicity with the surrounding lipid environment. It is obvious that exposure of polar residues to lipids will hamper proper integration of the corresponding stretch into the membrane, and will additionally destabilize secondary structure[22]. The location of polar residues at locations within the membrane is important, and this should be taken into account by selecting the fragment such that most of the polar residues find their interaction partner within the fragment if possible. On the other hand, looking at fragments also offers the unique possibility to study the behavior of an N-terminal part of the protein when it comes off the ribosome and exits the translocon in the absence of the remainder of the protein. We will come back to that point later on.

4.2 Synthetic aspects of biosynthesis of large fragments

Because of the ease of genetics, the relatively low costs of media and the possibility to introduce isotope labels in almost all flavors *E. coli* has generally been the host of choice for protein expression in NMR studies. In the case of GPCR fragments proteins have been usually derived from the insoluble form accumulated in inclusion bodies. Formation of inclusion bodies is believed to result from an overload of the intracellular chaperon machinery, or may be due to the presence of exposed hydrophobic sequences as in the case of GPCR fragments[23]. GPCR fragments are, by nature, prone to aggregation since they lack complementary TM helices and therefore expose surfaces to solvent or detergent that are usually buried within the helical bundle.

Purification of membrane proteins from inclusion bodies offers many advantages, e.g. the protein of interest often is already highly enriched in this compartment and accumulated to a much higher extent when compared to material that is directed into the inner membrane. Concerns arise about the refolding process, which for alpha helical membrane proteins is often a cumbersome task, and dedicated protocols have to be developed for each protein[24, 25]. Solubilization of inclusion

bodies mostly requires a denaturation step in a strong denaturant such as GdmCl or urea, followed by purification procedures as chromatographic separation, and a final refolding step.

The first often-encountered problem is to achieve protein overexpression – for NMR studies that means obtaining sufficient amounts of purified protein from 1-2 L because of the high costs of isotope-labeled media. For these reasons, expression as insoluble material into *E. coli* inclusion bodies has become the protocol of choice for most laboratories[22, 26-31].

As high and insoluble expression is in this case the ultimate goal, we would strongly recommend to use a vector under the control of the T7 promoter, as it has been shown that the T7 promoter, due to its inherent leaky control and high transcription rates, produces higher amounts of unfolded membrane proteins when compared to the L-arabinose promoter[32]. Moreover, the use of the T7 promoter is compatible with most of the commercial *E. coli* strains that can be tested for expression.

The first question to be solved is whether the GPCR fragment of interest can be expressed directly, without the use of any fusion partner. Direct expression, if achievable, if preferable as it allows to skip enzymatic or chemical removal of the fusion partner, often an inefficient purification step that can lead to loss of protein[33]. How to judge whether the GPCR fragment has been produced to sufficiently high levels? As milligram quantities are required to produce a final NMR sample, in our experience overexpression should be visible as a consistent band upon coomassie staining of a cell lysate sample. For the fragments of the Y4 receptor[26], direct expression could be achieved only for a limited portion of the receptor (e.g. for TM567), while mostly fusion to the TrpΔLE sequence expression was required, e.g. for TM67. While there is no single explanation why certain sequences are highly expressed and others are not, a body of evidence is suggesting that problems arise from unfavorable mRNA structures close to the translation initiation region, rather than codon bias[34].

The addition of short stretches of nucleotides close to the starting methionine has also been shown to greatly enhance membrane protein expression in a cell free system [35]. Moreover, we recommend to extensively optimize direct expression also by looking at single codon substitutions in the initiation region[36]. When direct expression cannot be achieved, the development of a good protocol using N-terminal fusions is often a trial-and-error procedure until high expression levels, stability, efficient proteolytic or chemical removal of the fusion, and homogeneous purification are achieved.

Inclusion bodies can be solubilized with the help of different agents, depending on the solubilization efficiency and the chosen purification strategy. Strategies for purifying membrane proteins from inclusion bodies are summarized in Figure 4.2:

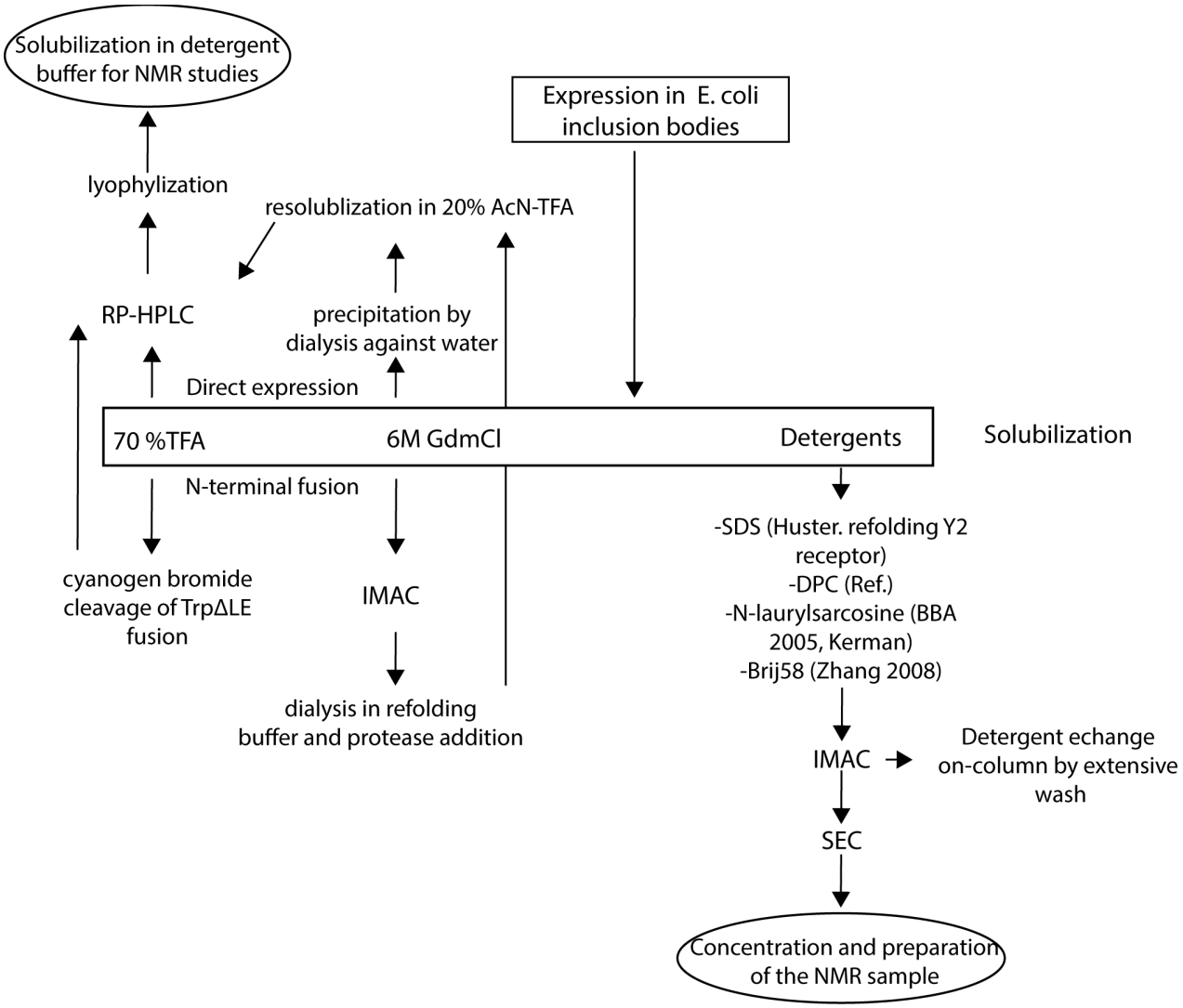


Figure 4.2: Schematic representation of the main strategies undertaken for the purification of GPCR fragments from *E. coli* inclusion bodies and subsequent refolding for NMR samples preparation.

We have successfully solubilized inclusion bodies directly in 70% trifluoroacetiacid (TFA), in which cleavage of the TrpΔLE sequence was achieved by addition of cyanogenbromide[27]. The reaction mix could be directly injected into a reverse-phase C3 column. This strategy allows purification in one-step, and was successfully employed by the Naider group for fragments from the Ste2p receptor. Unfortunately, it could not be applied to the Y4 receptor fragments, as inclusion bodies containing these proteins could not be solubilized by using only TFA. Instead, we used 6 M GdmCl[26], and after centrifugation the solubilized material was loaded on a Ni-NTA column. For

fusion proteins we modified the protocol such that the C3 protease was added to the GdmCl solution in Tris-buffer at pH 7.5. During removal of denaturant via dialysis, the C3 protease refolds and starts to cut the protein. When the GdmCl solution is lower than approximately 1 M, the liberated TM and TrpΔLE proteins precipitate. The precipitated material can be further resolubilized in few milliliters of 6M GdmCl in presence of high concentrations of reducing agents such as DTT or ®-mercaptoethanol (bME). The pH is lowered to ~2 and the mix injected into a reverse-phase C4 column. This protocol has been successfully used for expressing at high yields of triple-labeled TM67 from the Y4 receptor. The fragments TM12, TM45 and TM56 and TM567 could be directly expressed and investigated by NMR.

Unfortunately, not every fragment from the Y4 receptor could be efficiently solubilized after purification from inclusion bodies. While for TM12 very high expression levels were possible[37], expression of TM13 in amounts compatible with NMR studies was only successful after extensive optimization of the expression conditions (direct expression of the N-terminal His tagged protein in the *E. coli* Rosetta pLysS strain at 18˚C). Although the fragment could be purified under denaturing conditions using a C4-reverse phase column, we have encountered serious problems in solubilizing the fragments once it was precipitated or lyophilized, both in a variety of detergents or organic solvents. This prompted us to investigate expression into the inner membrane of *E. coli*, a method that may also be advantageous in producing correctly folded material. For fragments insertion into the membrane with the correct topology is an issue, and we decided to investigate this point in more detail.

4.3 Expression into the E. coli inner membrane and the topology of fragments

Expression of eukaryotic membrane proteins into the inner membrane of *E. coli* poses several challenges, and the mechanisms that determine topogenesis must be taken into account in order to produce well folded, and possibly functional material[21]. As eukaryotic membrane proteins often lack a signal sequence that can be recognized co-translationally by the bacterial SEC translocon, the introduction of an N-terminal bacterial signal sequence can be necessary to achieve correct export. In the case of GPCRs, the N-terminus needs to be localized into the periplasmic space of *E. coli* to ensure proper topology, and for this reason the maltose binding protein (MBP), with its cleavable signal sequence, has become the N-terminal fusion tag of choice for expressing functional GPCRs in *E. coli*[6, 38]. Insertion of misfolded proteins in the inner membrane of *E. coli* is mainly attributed to

failure of the bacterial translocon to correctly recognize determinants present on eukaryotic sequences, and to the fact that membrane protein folding is a slow process in comparison to the high rates of transcription achieved when using the T7 promoter. Geertsma *et al.* [31] demonstrated that overexpression of bacterial membrane proteins produces detectable amounts of unfolded material, which is higher when a T7 promoter was used instead of the more tightly regulated L-arabinose promoter.

Although the topology of G-protein coupled receptors is well known as several crystal structures are available[39], it is difficult to predict the organization of GPCR fragments. A full structural analysis using solution NMR will reveal details of membrane insertion, but the amount of work on a more comprehensive set of fragments is daunting. Topology analysis presents a useful alternative as it provides experimental low-resolution structural information on the number and organization of transmembrane helices.

Topology studies using reporters for cytoplasmic or periplasmic localization of the C-terminus were first established by von Heijne[40]. They used GFP as a reporter for cytoplasmic localization[41] and the alkaline phosphatase (PhoA) as a reporter for periplasmic localization. A schematic of the approach is presented in Figure 4.3.

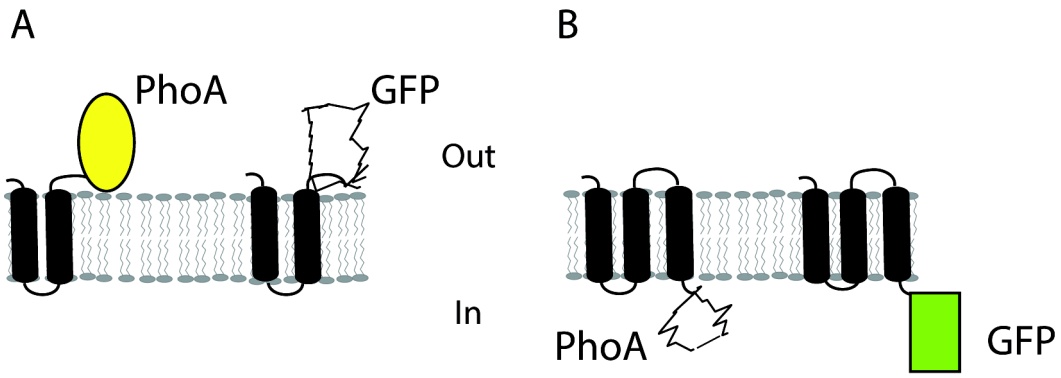


Figure 4.3: Reporter assay used to study the topology of membrane insertion

When the C-terminus is located in the periplasm PhoA activity can be measured (A). In case of cytoplasmic localization of the C-terminus, GFP fluorescence will occur (B).

Using this method, the group of Gunnar von Heijne found out that the *E. coli* ClcA H^+/Cl^- transporter[42] possesses two marginally hydrophobic alpha helices that were known to have a periplasmic C-terminus from the protein crystal structure, but topology analysis reported their C-terminus to be instead cytoplasmic. The correct localization could be restored by slightly increasing

the hydrophobicity inserting a single Leu residue in one of the alpha helices. These results highlight a possible mechanism happening during membrane protein insertion into the lipid bilayer, which would be otherwise unknown from the sole crystal structure: as the two alpha helices are marginally hydrophobic, the driving force for their correct insertion is given by the contacts with the other alpha helices that are already inserted into the membrane. It is in fact known that alpha helices are inserted co-translationally into the lipid bilayer either singularly, or as pairs if they can form early stabilizing interhelical interactions.

This example outline how topology studies can also give information on the folding process of membrane proteins. In an attempt to understand which fragments of the Y4 GPCR would acquire a unique topology when expressed into the membrane, we have performed a topology analysis by truncating the entire receptor by one helix at a time, starting from both the N-terminus and the C-terminus[43]. Since the N-terminal truncations were made in a way that the N-terminus was predicted to be localized either in the periplasmic or in the cytoplasmic compartments, depending on how many helices were removed from the N terminal end, we have used two different N-terminal fusion proteins to perform the study. The transmembrane region of the bacterial peptidase Lep was used to ensure proper periplasmic orientation[44], while Mistic was used to ensure cytoplasmic orientation to those fragments that started at truncations made in intracellular loops.

The information acquired by combining the results from the two reporters for cellular localization allowed quantifying the fraction of expressed protein that folds in a topology that resembles the organization in the entire receptor. We could observe that N-terminal fragments do not acquire a unique topology, however, the fraction displaying the correct localization of the C-terminus increases with the number of added helices. Interestingly, fragments truncated from the C-terminal end of the receptor possess a more defined topology. A scheme displaying the possible topologies for the two short fragments TM1 and TM12 of the Y4 receptor and experimental data on the location of the C-terminus are depicted in Figure 4.4.

In case of TM1 two topologies are possible for each fusion: a dual topology is observed in case of Lep (Figure 4.4A), while a predominant cytosolic location exists for mistic-TM1. Conversely, dual topology exists for mistic-TM1-TM2 and a predominant periplasmic localization of the C-terminus for Lep-TM1-TM2.

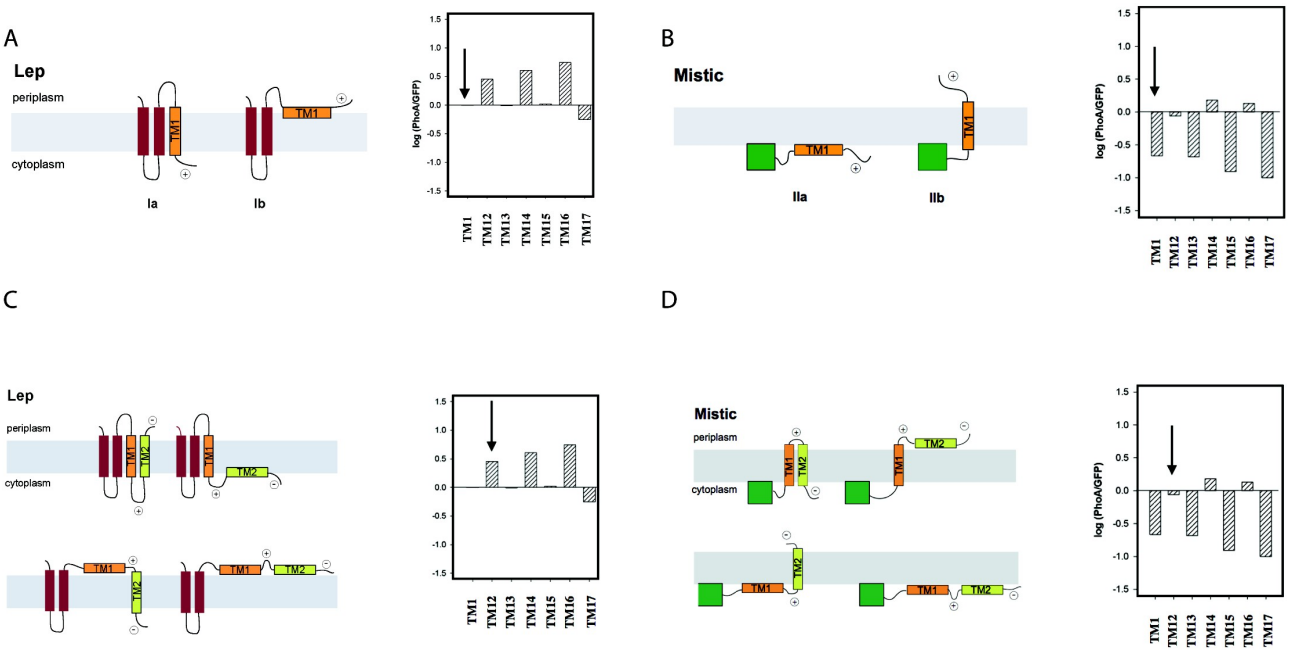


Figure 4.4: Topology studies of the Y4 receptor

Topologies for TM1 fused to Lep (A) or Mistic (B) or for TM12 fused to Lep (C) or Mistic (D). The corresponding relative strengths of PhoA activity or GFP signals are depicted on the right.

We believe that these data can be rationalized when considering the propensity of the TM helices for membrane insertion based on thermodynamic data for partitioning of the amino acids contained in the corresponding stretches[45] (Figure 4.5A), and the distribution of positive and negative charges in the loops (Figure 4.5B). In combination with our NMR studies on TM1, TM12 and TM123 (*vide infra*) the data suggest that the Y4 GPCR might be composed of a folding nucleus localized at the C-terminus and composed of helices TM5, 6 and 7, while the N-terminal side of the protein, characterized by an unusual high abundance of polar residues in the TM segments, can form a stable helical bundle only in presence of the C-terminus of the receptor[22].

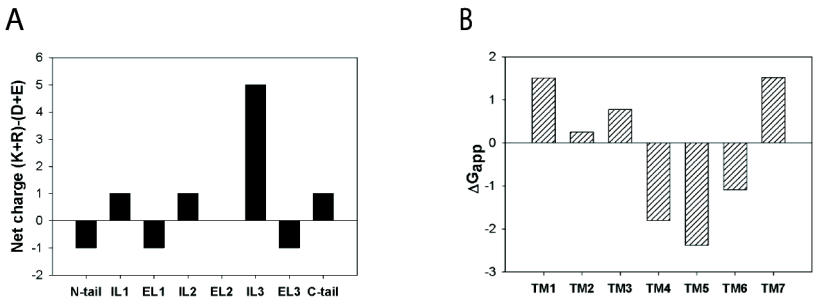


Figure 4.5: Net charge of loops of Y4 receptor (A) and free energies for transferring the corresponding TM stretches into the hydrophobic environment (B)

4.4 Conformational preferences of large fragments

The literature of membrane protein fragments studied by NMR has recently been reviewed[27]. Much of this work has been done in organic solvent mixtures, and we will not comment on that work here. Instead we would like to mention studies of GPCR fragments in detergent micelles.

Our group has studies large fragments from two GPCRs in detergent micelles: the work on the yeast Ste2p receptor was performed in collaboration with the groups of Naider and Becker from the US. Ste2p presents a class D yeast GPCR that is targeted by the α-factor and thereby prepares cells for mating. The class A (rhodopsin-like) human Y receptors bind to peptides from the neuropeptide Y (NPY) family of neurohormones and trigger many physiologically important functions involved in memory retention, food uptake etc. Four functional subtypes of the Y receptor have been identified in humans so far (Y1, Y2, Y4 and Y5)[46].

We initially investigated conformational preferences of the extracellular domains of all Y receptors[47]. They are largely unfolded, however, in case of the Y4 receptor, an amphiphilic helix anchored on the micelle surface is observed, connected via a long and flexible loop to TM1. Interestingly, PP binds with 50 μM affinity to that loop as determined by surface-plasmon resonance, largely via electrostatic interactions[48]. We previously postulated that ligands for the Y receptors in a first step associate with the membrane surface, and diffuse laterally[49-52]. In proximity a weak interaction of PP with the N-terminal domain helps transferring the peptide from the membrane-bound state to the receptor binding-pocket in a fly-casting mechanism (Figure 4.6).

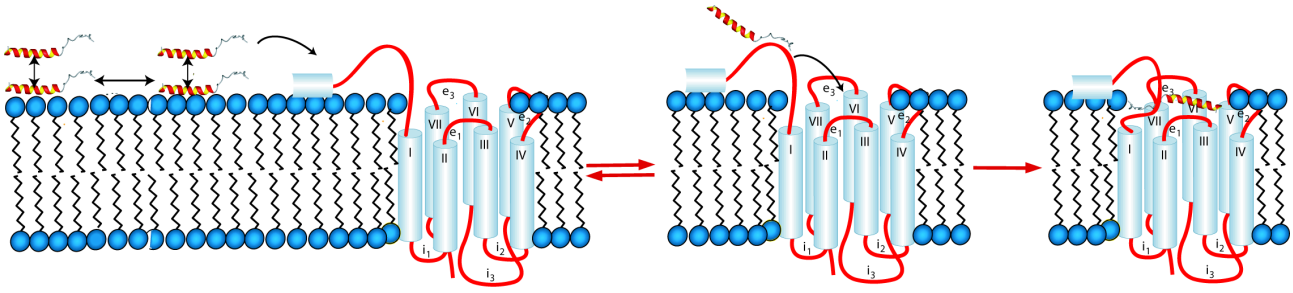


Figure 4.6: Model for binding of PP to the Y4 receptor

Model for binding of PP to Y4 receptor comprising association with the membrane surface (left), formation of transient contacts to the N-terminal domain (center) and diffusion into the orthosteric binding pocket (right)

The constructs were chosen to comprise one or more *entire* putatively helical stretches. Work on the Ste2p receptor was performed on TM7[53], TM1, TM12[54], TM123 and TM127 (in preparation,

see Chapter 3). In case of the Y4 receptor we studied TM12[22, 37], TM123, TM45, TM56, TM67[26] and TM567. Unfortunately, we never managed to obtain quantities of TM123 of the Y4 sufficient for NMR studies. In addition we investigated the extracellular N-terminal domains of all Y receptors[47, 55].

The biochemical aspects of this work are described above. Biophysical properties of the different proteins were quite variable, from pleasant to work with (stable at elevated temperatures, monomeric, resulting in peaks of homogenous line-widths) to difficult (aggregating in detergent, resulting in spectra with variable line-widths or lack of many signals, unstable over time as apparent by the formation of new peaks etc). We do not want to focus on the difficult technical aspects of this work but rather report on what we have learned from these studies and how that may help us to understand the underlying biology of these receptors.

Our work on structures of TM domains started with the structure of TM7, a construct that also comprises 40 residues from the C-terminal tail, from the Ste2p receptor[53]. The protein was expressed and purified in the lab of Fred Naider, and we determined the structure in DPC micelles. The NMR data revealed that TM7 is not stably formed, but rather exists in form of two helical stretches interrupted around a central Pro residue. Experiments with micelle-integrated spinlabels revealed that the helices are not stably anchored in the hydrophobic portion of the micelles but likely are only transiently buried. Likewise, we determined the structure of TM1 of the Ste2p receptor in LPPG/DPC mixed micelles. Again, two rather well-defined helices are observed, that are spaced by a GSRVG pentapeptide stretch (see Chapter 3).

The structure of TM12 of the Ste2p receptor, that contains 20 of the 50 residues from the extracellular N-terminal domain, the first two TM helices as well as residues from EL1, was determined by us in LPPG micelles[54]. Secondary structure was rather well defined, and this was supported by an almost complete set of Hα,Hβ i,i+3 NOEs in the helix-spanning regions. Interhelical contacts could be detected as interhelical NOEs between methyl groups. The number of restraints is still low and compared to a soluble protein the structure is still underdetermined with respect to experiment restraints but certainly rather well-defined for a helical membrane protein.

The corresponding structure of TM12 of the Y4 receptor did not reveal the presence of tertiary structure[21]. Structure prediction using chemical shifts indicated that secondary structure is formed,

however, interrupts occur close to Gly, Pro or polar residues. A comparison of these two structures is depicted in Figure 4.7.

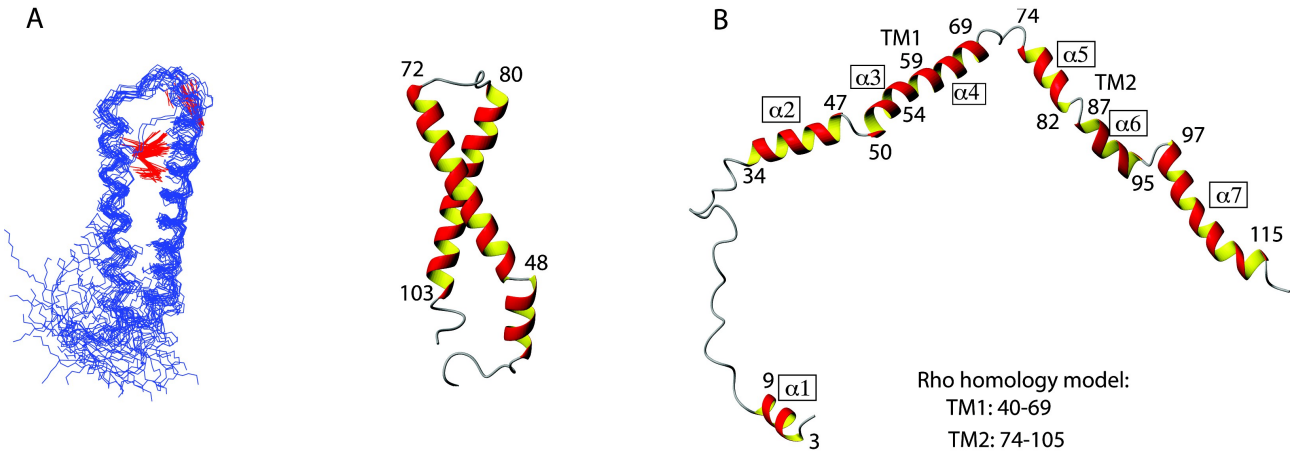


Figure 4.7: Structures of TM12.

A *20 lowest energy conformers (left) and single conformer (right) of TM12 from Ste2p.* ***B*** *Single conformer of TM12 from the Y4 receptor. Predicted regions of TM1 and TM2 based on the rhodopsin homology model are indicated at the bottom.*

In addition, we have investigated two fragments from the Ste2p receptor (TM123 and TM127), in which yet another TM helix is appended to the TM12 construct, thus systematically increasing the size of the fragments. While the addition of the second helix in TM12 appeared to have a structurally beneficial effect compared to TM1, effectively stabilizing the central GSRSVG hexapeptide stretch (unpublished data), elongation of the construct by the third TM helix to TM123 seems to increase its propensity to aggregate. The reason for this can be found in several polar residues located in the center of TM3, which presumably are missing their interhelical bonding partners. Paramagnetic relaxation enhancement NMR experiments using soluble spinlabels have indeed confirmed that the center of TM3 around an ETS sequence in the context of the TM123 construct is indeed solvent accessible and therefore not properly integrated into the micelle (see Chapter 3). This probably results in destabilized species, which are prone to aggregation.

In the TM127 fragment, the seventh TM helix is covalently linked to TM2, as it is proposed, based on a homology model, that most interhelical polar contacts are formed between TM1, TM2 and TM7. In contrast to TM123, TM127 does not show any tendency to aggregate. However, structure determination of TM127 has shown that it is a dynamic system, with a kink in TM7 around the Leu-Pro-Leu motif on the one hand, and with TM2 and TM7 competing for the same

unspecific interhelical contacts with TM1(see Chapter 3). This further confirms the hypothesis that tertiary structure in α-helical membrane proteins is determined predominantly by 'unspecific' interhelical interactions, which require only the presence of sequentially undefined α-helix, such as 'knob-in-the-hole' contacts[18, 19], in combination with several specific (mostly polar) interhelical contacts. Secondary structure is essentially defined by the amino acid sequence and its propensity to form helices. However, it appears to be critical that specific polar interhelical interactions are satisfied at all times, as can be seen by the comparison of the respective TM12 fragments of the Ste2p and Y4 receptor[22].

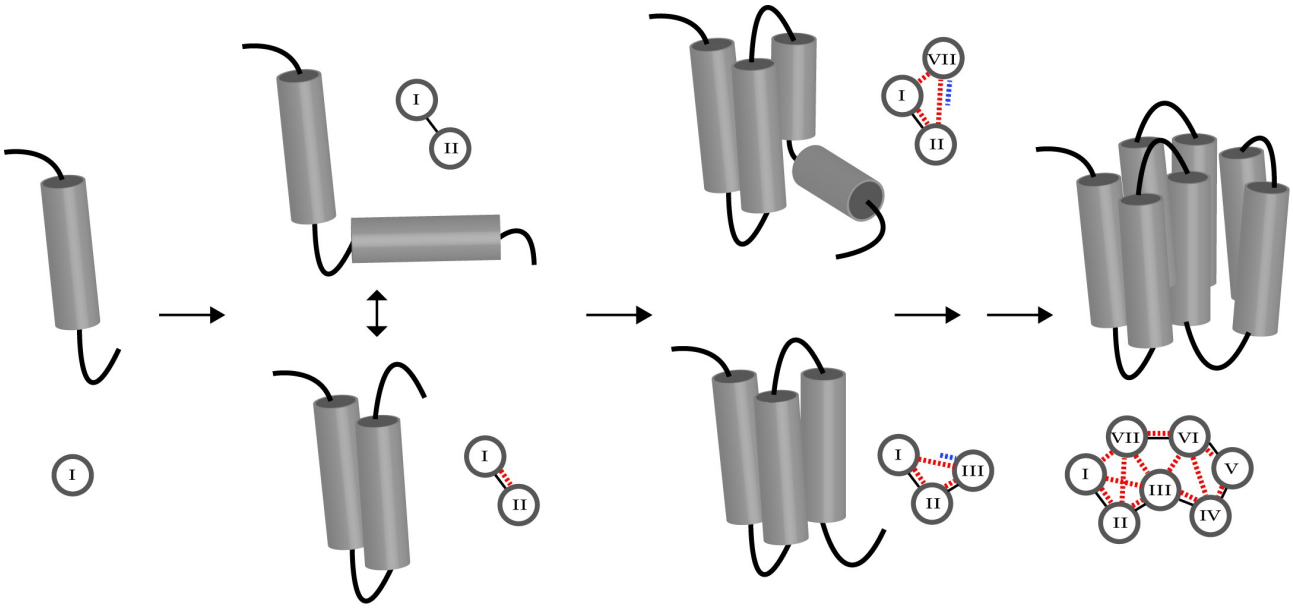


Figure 4.8: Schematic representation of the investigation of interhelical contacts by the systematic elongation of receptor fragments

interhelical contacts are indicated by red dotted lines. Potentially unsatisfied contacts are indicated blue dotted lines

Another important technical benefit of the study of systemically elongated membrane protein fragments is that it provides a tool for the resonance assignment through peak adaptations of overlapping fragments. We have observed that it is possible to transfer assignments from shorter onto longer fragments for overlapping sequences through comparison of strips from the ^{15}N-resolved 3D NOESY. Membrane proteins exhibit inherently few long-distance NOE peaks, which makes structural determination quite difficult, but also means that NOE information is mainly dependent on the local chemical environment, which is particularly true for amide protons. As the secondary structure of membrane proteins is predominantly determined by the amino-acid sequence, the ^{15}N-NOESY spectra of analogous residues in fragments can be compared efficiently.

We have managed to automate assignment transfer from shorter to longer fragments by comparing automatically picked NOESY spectra, based on manually picked [^{15}N,^{1}H]-HSQC spectra (see chapter 3). Automated assignment transfer from TM12 to TM127 resulted in over 90% correctly assigned backbone resonances. Indubitably, this process can also be expanded to larger fragments and subsequently possibly even whole GPCRs as long as the quality of the NOESY spectra is still acceptable. Typically, ^{15}N-NOESY experiments still exhibit reasonable spectra, even for large proteins. Furthermore, it might be possible to additionally transfer sidechain assignment, which would be of great help for structure determination and ligand binding studies of GPCRs.

4.5 Biophysical interpretation

From a thermodynamic perspective, partitioning of peptides or proteins into hydrophobic environments should promote folding in order to prevent the polar moieties of the peptide bond from forming unfavorable contacts with lipids. A vast amount of literature has reported about folding of peptides in the presence of detergent micelles that are unfolded in plain buffer[56].

Insertion of helices into hydrophobic environment obviously depends on biophysical properties of the amino acid sequence in the corresponding stretch. Wimley and White experimentally determined the free energy of transferring amino acids into the membrane interior or the membrane interface[45, 57]. Similarly, von Heijne has established an *in vivo* translocon-based scale that qualitatively largely agrees with the *in vitro* data[58, 59]. The work on TM1 and TM12 of Ste2p indicates that longer fragments better integrate into the membrane and more likely form tertiary structure. In addition, the topogenesis study of the series of Y4 receptor fragments clearly indicates that those fragments that contain more TM helices more stably insert. Importantly, TM1 is predicted to not partition into the membrane, and when taking this feature into account the topology of the truncated fragments can be predicted rather reliably (see Figures 4.4 and 4.5).

One particular issue with working with fragments is that polar residues that are located in central regions of the TM helices may not find their corresponding interaction partner and become exposed to the lipids. This is depicted in Figure 4.9C. Herein, the surface potential for a model of the Y4 receptor is depicted and compared to the surface potential of TM12. Clearly, the exposure of $Glu_{1.42}$ and $Asp_{2.50}$ in the center of the TM helices result in a charged surface patch that may perturb proper integration of the segment into the micelles. Our NMR studies on fragments indicate that

these polar residues result in instable membrane integration and possibly also destabilization of secondary structure.

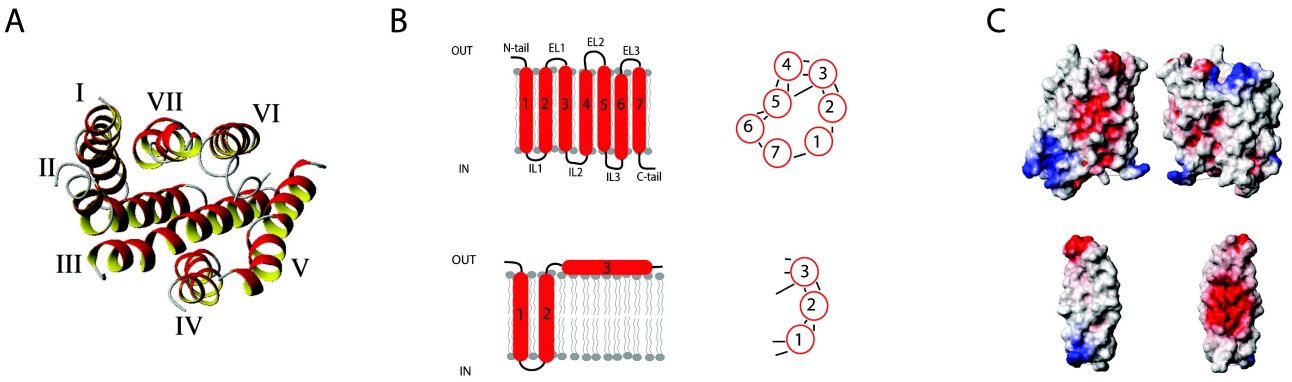


Figure 4.9: Contacts in entire receptors and fragments

A Top-view of a model of the Y4 receptor. B Topology of the entire (top row left) or TM13 (bottom row left) and their contacts (right). C Surface potential of the entire receptor (top) and of TM12 (bottom) from two different views. Note that for TM12 a native-like conformation was assumed for simplicity.

4.6 Mimics of GPCR loops

We have designed a loop mimetic in which the 3 extracellular loops of the Y4 receptor were grafted onto a protein that can be more easily biosynthesized and handled, resides in a membrane context and that can be characterized by solution NMR. To this end we grafted the loops onto a beta-barrel protein, the *E. coli* outer membrane protein A (OmpA) by replacing three of its four loops at one side of the barrel by the Y4 receptor native sequences[60, 61] (Figure 4.10):

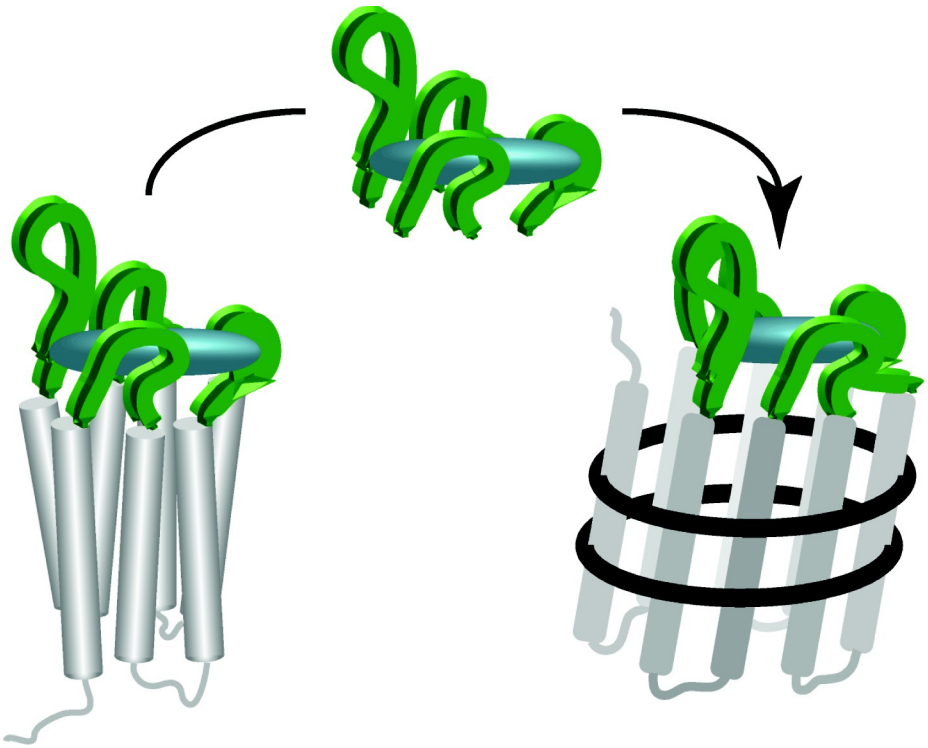


Figure 4.10: Grafting the extracellular loops of a GPCR onto a beta-barrel scaffold results in a chimeric receptor

The three loops can be transferred in various arrangements onto the scaffold. We decided to rank the different topomers according to how similar distances of loop-anchoring residues are with respect to known GPCR structures, and chose to biosynthesize 4 candidates that most closely resemble the native receptors. All of those expressed well and could be refolded. Using chemical shift mapping and saturation-transfer NMR techniques we could demonstrate that the chimeric receptor recognizes the native ligand NPY. We have also transferred the extracellular domain onto the remaining accessible beta-barrel terminus, but could not detect an increase in binding affinity[61].

4.7 Discussion

In this review we have described our efforts to study GPCRs from fragments. To summarize we have learnt that secondary structure in these fragments largely corresponds to the prediction from homology models. However, we have seen that tertiary contacts are mostly not formed, and the persistence of these contacts is related to the location and number of polar or charged residues in positions within the hydrophobic core of the membrane.

Furthermore, the investigation of systematically elongated fragments by NMR spectroscopy can provide detailed insights into early GPCR folding. The step-wise addition of helices to shorter fragments effectively simulates receptor biosynthesis, not only determining if interhelical contacts are formed, but also permitting the identification of specific contacts, which are required for proper folding, but also which are still unsatisfied. In addition, the important question can be tackeled whether these premature forms with unsatisfied polar residues already partition properly into the membrane or not. Furthermore, NMR spectroscopy allows the observation of multiple conformations and regions with high degrees of flexibility. It has to be noted though, that micelles as a membrane mimetic may have different biophysical properties compared to bilayer systems or even true biological membranes, which could alter the conformational preferences of these fragments. It is therefore imperative to compare the topology of fragments in micelles to their conformation in bilayers or even to the topology of the respective analogues expressed in native membranes.

In eukaryotes membrane protein biosynthesis of the polypeptide chain occurs at the ribosome in the ER starting at the N-terminus from which the nascent chain is transferred via the signal recognition particle (SRP), which targets the nascent chain to the ER inner membrane[62, 63]. From the SRP the chain is transferred to a large channel from the translocon machinery (Figure 4.11A). Inside the translocon hydrophobic segments are laterally gated through a cleft into the membrane interior[64]. Which segments are partitioned into the membrane interior is believed to be primarily decided based on hydrophobicity, a property that is reflected in the biophysical properties of the amino acid sequence[65].

The two-step model for helical membrane proteins introduced by Popot and Engelman proposes that TM helices individually insert into the membrane, diffuse in the membrane until they form the native contacts. Our data indicate that some of the TM helices of the Y4 receptor, but also from the Ste2p may not properly insert as entire helices when released into the hydrophobic core. Therefore, either bundling of TM segments occurs within the translocon (Figure 4.11B) or topological maturation once the polypeptide chain has been released into the hydrophobic interior (Figure 4.11C). Further experiments with more complex systems are required to analyze the folding pathway in more detail.

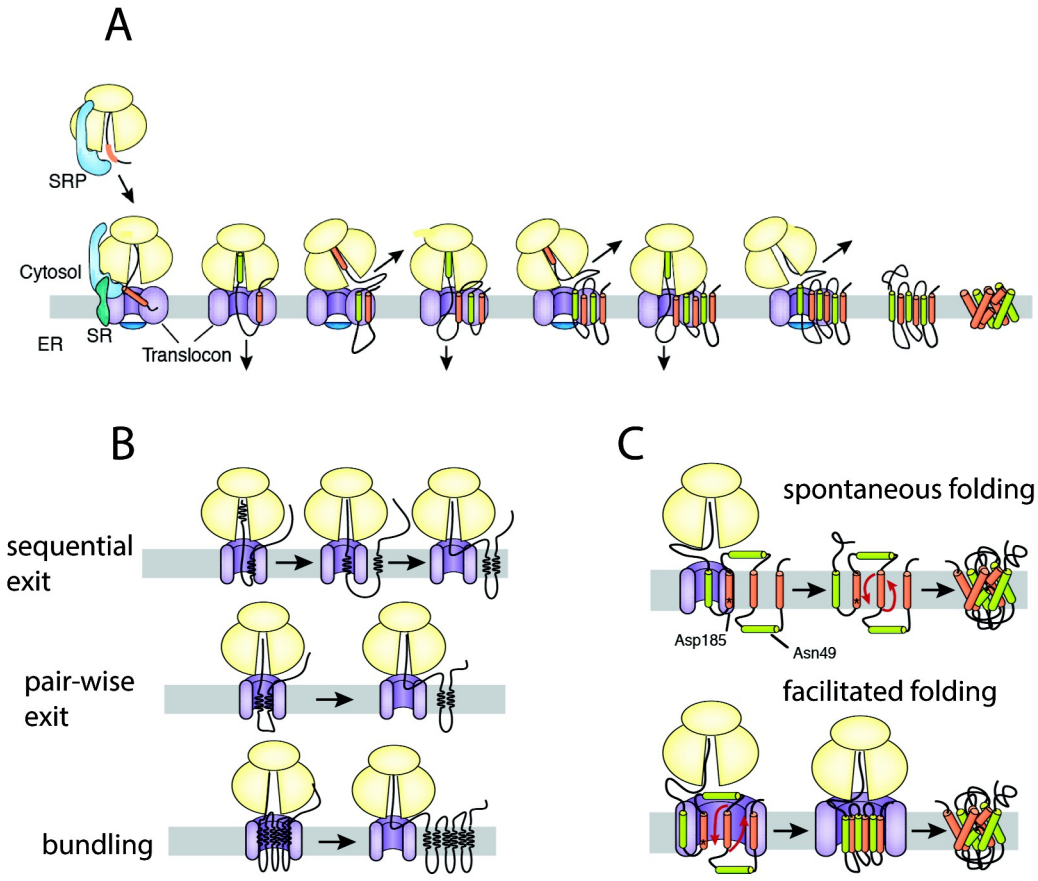


Figure 4.11: A Model for co-translational folding. B Various modes of formation of TM bundles. C spontaneous *vs*. facilitated folding as observed for aquaporin-4. Figure adapted from Skatch[62].

4.8 References

1. Schertler GF, Villa C, Henderson R. *Projection structure of rhodopsin.* Nature. 1993;**362:**770-2.
2. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science. 2000;**289:**739-45.
3. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, et al. *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science. 2007;**318:**1258-65.
4. Bockaert J. *[G-protein coupled receptors. Nobel Prize 2012 for chemistry to Robert J. Lefkowitz and Brian Kobilka].* Medecine sciences : M/S. 2012;**28:**1133-7.
5. Zerbe O. *First Solution Structures of Seven-Transmembrane Helical Proteins.* Angew Chem Int Ed Engl. 2011;**51:**860-1.
6. Egloff P, Hillenbrand M, Klenk C, Batyuk A, Heine P, Balada S, et al. *Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli.* Proc Natl Acad Sci U S A. 2014;**111:**E655-62.
7. Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. *Conformational thermostabilization of the 1-adrenergic receptor in a detergent-resistant form.* Proc Natl Acad Sci U S A. 2008**:** .
8. Scott D, Pluckthun A. *Direct molecular evolution of detergent-stable G protein-coupled receptors using polymer encapsulated cells.* J Mol Biol. 2013;**425:**662-77.
9. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D. *Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy.* Nat Struct Mol Biol. 2010;**17:**768-74.
10. Reckel S, Gottstein D, Stehle J, Löhr F, Verhoefen M, Takeda M, et al. *Solution NMR Structure of Proteorhodopsin.* Angew Chem Int Ed Engl. 2011;**50:**11942-6.
11. Popot JL, Engelman DM. *Helical membrane protein folding, stability, and evolution.* Annu Rev Biochem. 2000;**69:**881-922.
12. White SH, Wimley WC. *Membrane protein folding and stability: physical principles.* Annu Rev Biophys Biomol Struct. 1999;**28:**319-65.
13. Popot JL, Engelman DM. *Membrane protein folding and oligomerization: the two-stage model.* Biochemistry. 1990;**29:**4031-7.
14. MacKenzie K. *Folding and stability of alpha-helical integral membrane proteins.* Chem Rev. 2006;**106:**1931-77.
15. Langosch D, Arkin I. *Interaction and conformational dynamics of membrane-spanning protein helices.* Protein Sci. 2009;**18:**1343-58.
16. Mackenzie KR, Prestegard JH, Engelman DM. *A Transmembrane Helix Dimer - Structure and Implications.* Science. 1997;**276:**131-3.
17. Lemmon M, Flanagan J, Treutlein H, Zhang J, Engelman D. *Sequence specificity in the dimerization of transmembrane alpha-helices.* Biochemistry. 1992;**31:**12719-25.
18. Russ WP, Engelman DM. *The GxxxG motif: a framework for transmembrane helix-helix association.* J Mol Biol. 2000;**296:**911-9.

19. Langosch D, Heringa J. *Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils.* Proteins. 1998;**31:**150-9.

20. Dougherty DA. *Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp.* Science. 1996;**271:**163-8.

21. Bowie J. *Solving the membrane protein folding problem.* Nature. 2005;**438:**581-9.

22. Shao X, Zou C, Naider F, Zerbe O. *Comparison of fragments comprising the first two helices of the human Y4 and the yeast Ste2p G-protein-coupled receptors.* Biophys J. 2012;**103:**817-26.

23. Joergen B, Breitenstein A, Urlacher V, B5 uttner K, Lin H, Hecker M, et al. *Quality control of inclusion bodies in Escherichia coli.* Microbial cell factories. 2010;**9**.

24. Kiefer H. *In vitro folding of alpha-helical membrane proteins.* Biochim Biophys Acta. 2003;**1610:**57-62.

25. Booth P. *The trials and tribulations of membrane protein folding in vitro.* Biochim Biophys Acta. 2003;**1610:**51-6.

26. Kocherla H, Marino J, Shao X, Graf J, Zou C, Zerbe O. *Biosynthesis and spectroscopic characterization of 2-TM fragments encompassing the sequence of a human GPCR, the Y4 receptor.* ChemBioChem. 2012;**13:**818-28.

27. Cohen L, Fracchiolla K, Becker J, Naider F. *GPCR Structural Characterization: Using Fragments as Building Blocks to Determine a Complete Structure.* Biopolymers. 2014.

28. Xie X-Q, Zhao J, Zheng H. *Expression, purification, and isotope labeling of cannabinoid CB2 receptor fragment, CB2(180-233).* Protein Expression and Purification. 2004;**38:**61-8.

29. Zheng H, Zhao J, Sheng W, Xie XQ. *A transmembrane helix-bundle from G-protein coupled receptor CB2: biosynthesis, purification, and NMR characterization.* Biopolymers. 2006;**83:**46-61.

30. Zhang Y, Xie X-Q. *Biosynthesis, purification, and characterization of a cannabinoid receptor 2 fragment (CB2(271-326)).* Protein Expression and Purification. 2008;**59:**249-57.

31. Kerman A, Ananthanarayanan V. *Expression and spectroscopic characterization of a large fragment of the mu-opioid receptor.* Biochim Biophys Acta. 2005;**1747:**133-40.

32. Geertsma E, Groeneveld M, Slotboom D, Poolman B. *Quality control of overexpressed membrane proteins.* Proc Natl Acad Sci U S A. 2008;**105:**5722-7.

33. Vergis JM, Wiener MC. *The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal.* Protein Expression and Purification. 2011;**78:**139-42.

34. Kudla G, Murray AW, Tollervey D, Plotkin JB. *Coding-Sequence Determinants of Gene Expression in Escherichia coli.* Science (New York, NY). 2009;**324:**255-8.

35. Haberstock S, Roos C, Hoevels Y, Dötsch V, Schnapp G, Pautsch A, et al. *A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems.* Protein Expression and Purification. 2012;**82:**308-16.

36. Nørholm MHH, Toddo S, Virkki MTI, Light S, Heijne Gv, Daley DO. *Improved production of membrane proteins in Escherichia coli by selective codon substitutions.* FEBS LETTERS. 2013;**587:**2352-8.

37. Zou C, Naider F, Zerbe O. *Biosynthesis and NMR-studies of a double transmembrane domain*

from the Y4 receptor, a human GPCR. J Biomol NMR. 2008;**42:**257-69.

38. Tucker J, Grisshammer R. *Purification of a rat neurotensin receptor expressed in Escherichia coli.* The Biochemical journal. 1996;**317** 891-9.

39. Maeda S, Schertler G. *Production of GPCR and GPCR complexes for structure determination.* Current Opinion in Structural Biology. 2013;**23:**381-92.

40. Drew D, Sjostrand D, Nilsson J, Urbig T, Chin Cn, Gier JWd, et al. *Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis.* Proceedings of the National Academy of Sciences of the United States of America. 2002;**99:**2690-5.

41. Drew DE, Heijne Gv, Nordlund P, de Gier J-WL. *Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli.* FEBS LETTERS. 2001;**507:**220-4.

42. Cassel M, Seppälä S, Heijne Gv. *Confronting Fusion Protein-Based Membrane Protein Topology Mapping with Reality: The Escherichia coli ClcA H+/Cl− Exchange Transporter.* Journal of Molecular Biology. 2008;**381:**860-6.

43. Marino J, Geertsma ER, Zerbe O. *Topogenesis of heterologously expressed fragments of the human Y4 GPCR.* Biochimica et Biophysica Acta (BBA) - Biomembranes. 2012;**1818:**3055-63.

44. Bilgin N, Lee J, Zhu H, Dalbey R, von Heijne G. *Mapping of catalytically important domains in Escherichia coli leader peptidase.* EMBO J. 1990;**9:**2717-22.

45. Wimley WC, White SH. *Experimentally determined hydrophobicity scale for proteins at membrane interfaces.* Nature Struct Biol. 1996;**3:**842-8.

46. Berglund MM, Hipskind PA, Gehlert DR. *Recent developments in our understanding of the physiological role of PP-fold peptide receptor subtypes.* Exp Biol Med. 2003;**228:**217-44.

47. Zou C, Kumaran S, Walser R, Zerbe O. *Properties of the N-terminal domains from Y receptors probed by NMR spectroscopy.* J Pept Sci. 2009;**15:**184-91.

48. Zou C, Kumaran S, Markovic S, Walser R, Zerbe O. *Studies of the structure of the N-terminal domain from the Y4 receptor - a G protein-coupled receptor - and its interaction with hormones from the NPY family.* Chembiochem : a European journal of chemical biology. 2008;**9:**2276-84.

49. Bader R, Zerbe O. *Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state?* ChemBioChem. 2005;**6:**1520-34.

50. Bader R, Bettio A, Beck-Sickinger AG, Zerbe O. *Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection.* J Mol Biol. 2001;**305:**307-92.

51. Lerch M, Gafner V, Bader R, Christen B, Folkers G, Zerbe O. *Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles.* J Mol Biol. 2002;**322:**1117-33.

52. Lerch M, Mayrhofer M, Zerbe O. *Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors.* J Mol Biol. 2004;**339:**1153-68.

53. Neumoin A, Arshava B, Becker J, Zerbe O, Naider F. *NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor*

from Saccharomyces cerevisiae. Biophys J. 2007;**93:**467-82.

54. Neumoin A, Cohen L, Arshava B, Tantry S, Becker J, Zerbe O, et al. *Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles.* Biophys J. 2009;**96:**3187-96.

55. Zou C, Kumaran S, Markovic S, Walser R, Zerbe O. *Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family.* ChemBioChem. 2008;**9:**2276-84.

56. Bader R, Lerch M, Zerbe O. NMR of membrane-associated peptides and proteins. In: Zerbe O, editor. BioNMR in Drug Research. Weinheim: Wiley-VCH; 2002. p. 95-120.

57. White SH, Wimley WC. *Hydrophobic interactions of peptides with membrane interfaces.* Biochim Biophys Acta. 1998;**1376:**339-52.

58. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, et al. *Recognition of transmembrane helices by the endoplasmic reticulum translocon.* Nature 2005;**433:**377-81.

59. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, et al. *Molecular code for transmembrane-helix recognition by the Sec61 translocon.* Nature. 2007;**450:**1026-30.

60. Walser R, Kleinschmidt JH, Zerbe O. *A chimeric GPCR model mimicking the ligand binding site of the human Y1 receptor studied by NMR.* ChemBioChem. 2011;**12:**1690-3.

61. Walser R, Kleinschmidt JH, Skerra A, Zerbe O. *β-Barrel scaffolds for the grafting of extracellular loops from G-protein-coupled receptors.* Biological chemistry. 2012;**393:**1341-55

62. Skach W. *Cellular mechanisms of membrane protein folding.* Nat Struct Mol Biol. 2009;**16:**606-12.

63. von Heijne G. *Recent advances in the understanding of membrane protein assembly and structure.* Q Rev Biophys. 1999;**32:**285-307.

64. Osborne A, Rapoport T, van den Berg B. *Protein translocation by the Sec61/SecY channel.* Annu Rev Cell Dev Biol. 2005;**21:**529-50.

65. White S, von Heijne G. *How translocons select transmembrane helices.* Annu Rev Biophys. 2008;**37:**23-42.

Curriculum Vitae:

Name:	Martin POMS
Birth:	February 10th 1984, Wolfsberg, Austria
Nationality:	Austrian
Professional Address:	University of Zurich Institute of Organic Chemistry Winterthurerstrasse 190 CH-8057 Zurich, Switzerland

Education:

School:

09/1994 – 06/2002	*Stiftsgymnasium St. Paul im Lavanttal, Kaernten, Austria* Graduated 'summa cum laude'
08/2000 – 06/2001	Student Exchange *Canyon High School, Anaheim, California, USA*

Higher Education:

10/2002 – 02/2006	Bachelor Studies of Technical Chemistry *Graz University of Technology, Graz, Austria*
08/2006-12/2006	ERASMUS exchange *Corvinus University, Budapest, Hungary*
08/2006-01/2009	Master Studies - Biotechnology, Biochemistry and Food Technology *Graz University of Technology, Graz, Austria*
01/2008-07/2008	Diploma Thesis ' Backbone cyclization of the conotoxin ρ-TIA' *University of Queensland, Brisbane, Australia*
01/2010 – present	PhD studies in the group of Dr. Oliver Zerbe Life-science graduate school Zurich, Molecular Life Sciences program *University of Zurich, Zurich, Switzerland*

Relevant Work Experience:

01/2010 – present	PhD studies in the group of Dr. Oliver Zerbe *University of Zurich, Zurich, Switzerland*
08/2005 – 09/2005	Lab assistent *Verbund GmbH, St. Andrä i. L., Austria*
08/2002 – 09/2002	Research assistant *HOS Technik, St. Stefan i. L., Austria*

Extracurricular activities:

06/2010 – 09/2013	Representative of PhD students and Postdocs in the Institute of Organic Chemistry, *University of Zurich*
02/2011 – 09/2013	Representative of academic mid-level faculty in the Faculty of Science, *University of Zurich*
09/2012 –02/ 2013	Working Group 'PhD regulationsfor the Faculty of Science' *University of Zurich*

Further Education:

09/2010	Scientific writing Institute for Molecular Life Science, *University of Zurich*
05/2011	Ethics in Science Institute of Biomedical Ethics, *University of Zurich*
11/2013	Wissenschaft öffentlich kommunizieren (Communicating Science publicly) Zentrum für Weiterbildung, *University of Zurich*
10/2014	Business Know-How, Zentrum für Weiterbildung, *University of Zurich*

Grants and Scholarships:

08/2006	ERASMUS scholarship, *Graz University of Technology*
01/2008	Scholarship for short time academic research and expert courses abroad (KUWI), *Graz University of Technology*
08/2012	Travel grant of the Molecular Life Sciences program, *University of Zurich*
06/2013	Travel grant of the Molecular Life Sciences program, *University of Zurich*

Poster Presentations:

09/2010	"Structural Studies of the GPCR-fragment TM1-TM2-TM3 of Ste2p" Swiss Chemical Society Fall Meeting; *ETH Zurich, Switzerland*
08/2012	"NMR studies of GPCR fragments of the Ste2p receptor" International Conference on Magnetic Resonance in Biological Systems; *Lyon, France*
06/2013	"NMR studies of GPCR fragments" International School of Biological Magnetic Resonance; Future of Biophysics; *Erice, Italy*

Language skills:

German	native
English	fluent